THE VALIDATION OF A STUDENT SURVEY ON TEACHER PRACTICE

By

Ryan Thomas Balch

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements for

the degree of

DOCTOR OF PHILOSOPHY

in

Leadership and Policy Studies

August, 2012

Nashville, Tennessee

Approved:

Professor David S. Cordray

Professor Matthew G. Springer

Professor Mimi Engel

Professor Mark Berends

Though there is widespread evidence that teachers matter, a more challenging problem exists in attempting to measure teacher effectiveness. It can be argued that student feedback is an important consideration in any teacher evaluation system as students have the most contact with teachers and are the direct consumers of a teacher's service. The current paper outlines the development and preliminary validation of a student survey on teacher practice. Using data from a large-scale pilot in Georgia, the analysis finds that teacher scores on a student survey have a positive and marginally significant relationship to value-added estimates of teacher effects on student achievement. Further, there is a strong link between teacher scores and measures of academic student engagement and student self-efficacy. Finally, the paper investigates policy related issues that are pertinent to implementing student surveys as a component of teacher evaluation.

ACKNOWLEDGEMENTS

I would like to those who directly supported my doctoral work. I especially would like to thank Kathleen Mathers and the Governor's Office of Student Achievement for facilitating the student survey pilot in Georgia. Further, I would also like to thank all of the teachers and students who participated in the student survey pilot. Finally, I would like to thank the members of my dissertation committee for their continued support, guidance, and valued feedback. In particular, David Cordray provided essential assistance in learning about the survey development process through independent studies and serving as my committee chair.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1:

Introduction

Teacher evaluation plays a central role in today's education policy debates at the national, state, and local levels. In the past, teachers have primarily been evaluated by their principals based on classroom observations and other sources of evidence about teachers' practices. However, there is a growing recognition of the wide variation in teachers' effectiveness, combined with evidence that traditional forms of evaluation have failed to distinguish effective from ineffective teachers (Jacob & Lefgren, 2005) that has resulted in policymakers and some stakeholder groups promoting alternative means of evaluating teachers. In an era of increased accountability for both teachers and schools, determinations of effective teaching are only as valid as the instruments of evaluation.

Though we still seek an agreed upon definition of effective teaching, one consistent finding in education research is that there is significant variation among teachers in their ability to increase student achievement. Hanushek & Rivkin (2006) find that teachers near the top end of the quality distribution can lead their students to a full year's worth of learning more than teachers near the bottom end. Specifically, teachers in the 95th percentile had student gains of 1.5 grade level equivalents on the Texas Assessment of Academic Skills (TAAS) while teachers in the 5th percentile only had an increase of 0.5 during one school year. Even using a more conservative estimate, others have found that moving from an average teacher to one in the 85th percentile can result in a 4 percentile average test score gain for students in that teacher's class (Rivkin, Hanushek, & Kain, 2005). In the same data, this was roughly equivalent to the effect of reducing class size by 10 students.

Research has demonstrated that these substantial differences in achievement are both additive and cumulative for students. Having an effective teacher for three sequential years resulted in a 50 percentile point difference in achievement compared to students who experienced an ineffective teacher for three years (Sanders & Rivers, 1996). Moreover, subsequent teachers also appear unable to completely reverse either the positive or negative effects of previous teachers.

While there is evidence of differences among teachers, it is still challenging to quantify and measure this variation. Researchers and practitioners have sought to meet this challenge using several different methods of teacher evaluation ranging from observational evaluation rubrics and teacher portfolios to value-added calculations of teacher's contributions to student achievement. A less common method uses feedback from students to measure teacher quality, though there is no instrument that has been routinely employed in schools. To assist in incorporating the student perspective, the current investigation outlines the development and validation of an instrument to measure teacher effectiveness using student feedback.

When considering possible measures of teacher effectiveness in K-12 education, it can be argued that student perceptions of a teacher are an important consideration in any teacher evaluation system as students have the most contact with teachers and are the direct consumers of a teacher's service (Goe, Bell, & Little, Approaches to Evaluating Teacher Effectiveness: A Research Synthesis, 2008). Further, other measures such as value added estimates are not feasible for many teachers because they teach subjects or grade levels that do not have standardized tests. Without secondary measures such as student evaluations, these teachers may end up being judged by the performance of the school rather than receiving any individual feedback.

Even teachers for whom value-added measures can be calculated may benefit from the additional information provided by student evaluations, as the latter can address practices and outcomes that are not captured by standardized achievement tests but that might be important intermediate outcomes that will ultimately improve student learning (e.g., teachers' effectiveness at promoting student interest in the subject, the frequency and quality of feedback teachers provide to students). Student surveys that provide information on specific practices can form the basis of targeted professional development in the areas where teachers obtain low ratings. Furthermore, combining student evaluations with other forms of assessment can prevent manipulation and gaming behavior in a high stakes environment.

A final benefit of student surveys relative to observational evaluations is that student surveys have the potential to provide similar information at a fraction of the cost and time. If one assumes that each teacher in a school building is observed four times per year for 30 minutes along with 30 minutes for pre and/or post-observation conferences and time spent writing reports, each teacher requires roughly 5-6 hours of time from a supervisor or lead teacher in a given school year. In a large district with 150 schools and 15,000 teachers this can translate to the full-time salary of roughly 40 employees at a total cost (salary plus benefits) of more than $4.0 million before even considering the cost of training and licensing fees. With student surveys requiring minimal staff, the reduced cost could potentially provide high quality feedback at a fraction of the cost of other measures.

Understanding how student ratings relate to education outcomes is of great importance in the current policy environment. In their applications for Race to the Top, five states indicated that student feedback would be a part of teacher evaluation systems (Learning Point Associates, 2010). Further, the state of Georgia's successful Race to the Top application noted that student feedback would potentially count for 10% of a teacher's evaluation in tested subjects, and 40%

of a teacher's evaluation in non-tested subjects (Georgia Department of Education, 2010). Given the policy relevance of student surveys and the potential role that student surveys play in a teacher's evaluation, it is essential that states and districts implement an instrument that has undergone proper validation.

The survey validation described below follows a framework that seeks to establish construct validity through multiple sources of evidence. The main construct of interest is that of effective teaching as defined by teacher behaviors. Construct validity refers to whether a scale measures the underlying construct that it is intended to measure (Messick, 1989). Evidence for construct validity comes from content validity, convergent validity, and predictive validity (a more thorough description of the validation framework is presented in the methods section of this paper). Content validity ensures a measure has adequate coverage of the content it seeks to measure and will be established by drawing upon effective teaching practices from the literature as well as common observation rubrics. Convergent validity compares scores on a measure to other instruments that intend to measure a similar concept. Convergent validity will be established with measures of academic student engagement and academic self-efficacy. Finally, predictive validity determines whether scores on a measure can predict future scores on measures of similar constructs. This analysis will investigate whether there is a relationship between teacher scores on the survey and a teacher's average value-added. To test both concurrent and predictive validity, a large-scale pilot was conducted in the spring of 2011 in seven districts as part of Georgia's Race to the Top initiative.

Value-added, academic student engagement, and self-efficacy were chosen based on their relationship to important education outcomes. Having high value-added teachers has been associated with greater future income and college attendance (Chetty, Friedman, & Rockoff, 2011). Further, academic self-efficacy has been linked to adaptive patterns of learning (Midgley

et al., 2004). Finally, student engagement is positively related to achievement on standardized tests and improved grades and negatively related to outcomes such as dropping out of school (Fredricks & McColskey, 2011).

Overall, the investigation will be guided by the following research questions:

- What effective teaching practices have been identified through the literature on teacher practice and validated observation rubrics?
- What is the relationship between a teacher's total score on a student feedback survey and estimates of a teacher's value-added to student achievement?
- What is the relationship between a teacher's total score on a student survey and other outcomes such as academic student engagement and academic self-efficacy?
- How have teachers used student feedback to inform instruction?

The next section reviews the literature on teacher quality and teacher evaluation. Following this is a description of the survey development process and an outline of the validation methods including cognitive interviews and pilot testing. Next, the paper presents the results of the pilot project in Georgia including the relationship between teacher scores on the survey and student engagement, self-efficacy, value-added, as well as internal properties of the survey. Finally, the paper concludes with policy considerations as well as recommendations and directions for future research.

Chapter 2:

Review of Literature

Teacher Quality

A precursor to developing measures of teacher quality is agreeing upon a definition. Complicating the matter is the fact that teacher quality can be defined in a number of ways. These may include teacher qualifications, teacher content knowledge, teacher characteristics, or actual teaching behaviors, with each showing a different relationship to student achievement.

Teacher qualifications include aspects such as teaching experience, advanced degrees, certification, and subject matter knowledge. Teacher experience predicts student achievement in some studies (Clotfelter, Ladd, & Vigdor, 2006; Harris & Sass, 2007) , but often the effect is limited to the first few years of a teacher's career (Hanushek, Kain, O'Brien, & Rivkin, 2005; Rockoff J. , 2004). For level of education, research consistently fails to find a relationship between advanced degrees and student achievement (Harris & Sass, 2007; Hanushek, Kain, O'Brien, & Rivkin, 2005; Clotfelter, Ladd, & Vigdor, 2006). Overall, Goldhaber (2002) finds that only 3 percent of a teacher's contribution to student learning was associated with teacher experience, degree attained, or other observable characteristics. These results call into question the fact that the vast majority of district salary structures reward teachers for qualifications – advanced degrees and years of teaching experience – that bear little relationship to student outcomes.

A more promising measure of teacher quality is teacher content knowledge. Most studies investigating content knowledge use teacher certification scores as a proxy; with results generally showing a positive relationship (Greenwald, Hedges, & Laine, 1996; Rowen, Correnti,

& Miller, 2002; Ferguson, 1991). These, however, are general measures of content that do not inform the types of knowledge or ability that a teacher requires to be an effective teacher. One study looked specifically at performance on instruments designed to test a teacher's mathematical knowledge for teaching and found a statistically significant ($p<.05$) and positive relationship to gains in student achievement (Hill, Rowan, & Ball, 2005). Although there appears to be evidence of a link between content knowledge and achievement, the type of content knowledge that is assessed is dependent on the instrument or measure.

Next, a large body of research has investigated what teacher characteristics are most associated with increased student achievement, with no clear consensus that any measured characteristics have an impact (Goe, 2007; Rockoff, Jacob, Kane, & Staiger, 2008; Goldhaber D. , The Mystery of Good Teaching, 2002). Characteristics such as race, ethnicity, and gender do not have a significant relationship to student achievement (Ehrenberg, Goldhaber, & Brewer, 1995), but there is evidence from the STAR randomized class size experiment that students with teachers of the same race have increased achievement in both reading and math (Dee, 2004). Rockoff et al. (2008) investigated a range of non-traditional teacher characteristics including content knowledge, cognitive ability, personality traits, and feelings of self-efficacy. They find that very few of these predictors have a significant relationship to achievement when analyzed individually, but factors that combine cognitive and non-cognitive teacher skills have a modest relationship (Rockoff, Jacob, Kane, & Staiger, 2008).

Finally, teacher quality may be defined by actual teacher behaviors that are associated with increased student achievement. Beginning in the 1960's and 1970's, there was a push to determine what teacher practices were associated with increased student achievement (Schacter & Thum, 2004). For instance, certain studies may look at specific practices such as the use of group work or stating the lesson objective at the beginning of the class. A number of reviews

have consolidated findings from these individual studies in an effort to present behaviors that show consistent relationships to student achievement. The categories from major reviews are shown in Appendix A.

While there are differences, a considerable amount of overlap exists among these reviews. For instance, providing high quality academic feedback is noted in several reviews as having an association with higher student achievement. Schachter and Thum (2004) call for "frequent, elaborate, and high quality academic feedback", Good and Brophy (1986) note the importance of "monitoring students' understanding, providing feedback, and giving praise", Emmer and Evertson (1994) emphasize that "all student work, including seatwork, homework, and papers, is corrected, errors are discussed, and feedback is provided promptly", and Marzano (2001) outlines a large body of research indicating the importance of teachers providing feedback. Other categories that overlap among the reviews include clarity of presentation, managing behavior promptly, reinforcing student effort, and having appropriate pacing. Next , we turn to measuring these behaviors.

Measures of Teacher Behaviors

The knowledge of what teacher behaviors may be associated with student achievement is most relevant if these practices can be measured. This is especially true when one considers information asymmetry from the principal-agent framework. Principal-agent theory describes the situation in which an employer (principal) hires an employee (agent) to perform a task (Alchian & Demsetz, 1972). The problem arises when the agent receives the same compensation regardless of the quality of work or effort level, sometimes leading to a reduction in both (Eisenhardt, 1989). The principal-agent problem is especially relevant in situations where there is not a clearly defined output of performance and low levels of supervisor monitoring, situations

that occur frequently in the teaching profession. When employees have lower incentives for increasing effort, it is argued that it is more efficient to replace fixed wages with compensation that links employees' pay to their performance (Alchian & Demsetz, 1972). In this regard, the type of incentives and method of measurement impact ways that systems address the inherent problem that principal-agent theory outlines.

This issue is particularly relevant in the field of education due to the lack of specificity regarding what product should be produced. As a result of this and wage discrepancies in the past, nearly all K-12 teachers in the United States are paid according to a single salary schedule that rewards advanced degrees and years of experience (Podgursky & Springer, 2007). The ability to measure teacher behaviors in a valid and reliable fashion has the potential to reduce this information asymmetry and provide a basis for more strategic compensation. There are a variety of instruments and techniques that have been implemented to measure teacher behaviors that will now be discussed.

Self-Evaluation

The first form of measuring teacher behaviors is to have teachers assess their own practices. Two examples of this technique are teacher surveys and logs. Many national surveys ask teachers about practices used during the entire year such as the Early Childhood Longitudinal Study, the National Assessment of Educational Progress, or the Schools and Staffing Survey's Teacher Follow-Up Survey. While these surveys tap a nationally representative population, they require that teachers make assessments of their practice from the entire year and may be subject to teachers responding with socially desirable answers or error due to problems with remembering (Rowan, Jacob, & Correnti, 2009). An alternative to large surveys is the use of instructional logs, a process by which teachers document content coverage and teaching

strategies on a more regular basis. While not as accurate as independent observation because of bias due to self-report, instructional logs have been found to be valid, reliable, and cost-effective (Rowan, Jacob, & Correnti, 2009). To establish validity, teacher logs were compared to researcher evaluation, finding that teacher-observer match rates ranged from 73 to 90 percent. Though self-evaluation may be useful for research or documentation of instructional practice, it is unlikely that this alignment would persist in a high-stakes environment.

Analysis of Classroom Artifacts and Portfolios

A second possible method for evaluation includes the analysis of classroom artifacts such as lesson plans, teacher assignments, assessments, scoring rubrics, and student work. While many systems use some sort of artifact analysis, a structured and valid protocol for evaluation is essential. Examples of such protocols include the Instructional Quality Assessment done by the National Center for Research on Evaluation, Standards, and Student Testing (Matsumura, Slater, Junker, et al., 2006) and the Intellectual Demand Assignment Protocol (IDAP) developed by the Consortium on Chicago School Research (Newmann et al., 2001). The IDAP showed both high inter-rater reliability (90 percent agreement) and that students of teachers that scored high on the instrument had learning gains on the Iowa Test of Basic Skills that were 20 percent higher than the national average. Though some findings indicate that teacher ratings using these artifacts are correlated with outcomes, there is a lack of research conducted by independent researchers (Goe, Bell, & Little, Approaches to Evaluating Teacher Effectiveness: A Research Synthesis, 2008).

Portfolios are a further option that may include similar teaching artifacts, yet in this case teachers prepare their own samples. It also may include other evidence such as statements of teaching philosophy or videotaped lessons. A common example of this is National Board for

Professional Teaching Standards Certification, which research indicates is successful in identifying high-quality teachers even though the process of certification may not improve effectiveness (Hakel, Koenig, & Elliot, 2008). It also remains to be seen whether this type of evaluation would be practical in a high-stakes setting.

Classroom Observation

Classroom observations represent one of the most commonly used evaluation systems for teachers (Goe, 2008). There are countless variations in frequency, instrument, rating scales, and protocol. Some of the main issues to consider with observations are the validity of the instrument and the reliability of rating, particularly if ratings are attached to financial rewards or job security. Examples of instruments that have been validated for their relationship to student achievement include Charlotte Danielson's (1996) *Framework for Teaching* and the Classroom Assessment Scoring System (CLASS) for grades K-5 (Pianta, La Paro, & Hamre, 2006). Outside researchers found that a student with a teacher in the top quartile (according to Danielson's rubric) would score 0.10 standard deviations higher in math and 0.125 standard deviations higher in reading than a student assigned to a teacher in the bottom quartile (Kane, Taylor, Tyler, & Wooten, 2010). Further, a project funded by the Gates Foundation investigated the relationship between several measures of teaching and value-added estimates of student achievement. Teacher ratings from the Danielson Framework had a 0.19 correlation with student achievement in math and a 0.11 correlation with student achievement in ELA. For CLASS, the correlations were 0.24 and 0.10 respectively. Though small, these correlations were all statistically significant ($p < .05$).

Further, observational rubrics can also consolidate research on teacher behaviors that are associated with increased student achievement. For example, Schacter and Thum (2004)

developed six teaching standards of teacher behavior in the areas of questions, feedback, presentation, lesson structure and pacing, lesson objectives, and classroom environment. The categories were based on teaching models that combined effective practices and garnered large effect sizes (d = 0.46 – 1.53) in reading, language, mathematics, and social science for teachers that were randomly assigned to training with these models (Gage & Needles, 1989). These were combined with five standards of teaching strategies found to show increased student achievement that included grouping students, encouraging student thinking, providing meaningful activities, motivating students, and teacher knowledge of students. All together Schacter and Thum developed a rubric with 12 different teaching standards and a corresponding rubric to determine teacher quality. The rubric was tested with 52 elementary school teachers and the authors found that students of teachers who implement the practices in these 12 categories make considerable gains (standardized regression coefficient of 0.91) in achievement (Schacter & Thum, 2004).

While observations using these rubrics have demonstrated a link to student achievement, the investment necessary for complete implementation is large. Extensive training is necessary for all evaluators, and immediate connection to incentives may increase the potential for errors due to inexperience with an evaluation instrument. Though observational evaluation is the most common form of personnel evaluation for teachers, the rubrics many school systems employ do not require the training or expertise necessary for more advanced instruments.

Finally, the observations rubrics described above represent the most promising instruments that have gone through a sound validation process. In reality, the rubrics employed by the majority districts fail to differentiate among teachers with one study demonstrating that over 99% of teachers were judged as highly effective according to the district's observation rubric (Weisberg, Sexton, Mulhern, & Keeling., 2009).

Student Feedback – Higher Education

The majority of empirical investigations of student feedback in teacher evaluation have occurred within higher education. From these studies, student ratings appear to be both reliable and valid measures of instructional quality. Student ratings of college professors in subsequent years have correlations between .87 and .89, suggesting they are stable and reliable (Aleamoni, 1999). Further, a meta-analysis of studies on student ratings found an average correlation of .43 between mean student ratings of instructors and mean student performance on common final exams in multi-section courses. This is combined with positive correlations between student feedback and ratings from colleagues and external observers (Renaud & Murray, 2005). As Renaud and Murray (2005) note in their review of the literature, "the weight of evidence from research is that student ratings of teacher effectiveness validly reflect the skill or effectiveness of the instructor" (p. 930).

Despite these findings, it is possible that extraneous factors could bias these ratings. Some have found a negative relationship between student ratings and expected course grades, indicating that students rate challenging teachers lower (Rodin & Rodin, 1972). In an extensive review, it was found that twenty-four studies found zero relationship, and thirty-seven studies found contradictory results to this notion (Aleamoni, 1999). Based on this evidence, it appears that students are able to separate effective instruction from their own personal academic expectations.

Finally, an instructor's personality characteristics could influence students' perception of their effectiveness, resulting in positively biased ratings. A review of the literature finds seventeen studies that find students are "discriminating judges of instructional effectiveness," (Aleamoni, 1999, p. 154) with students showing considerable ability to differentially rate instructors in categories such as method of instruction, course content, general course attitude,

and interest and attention. While some overlap between an instructor's personality characteristics and their effectiveness should be expected, it is important to document that student ratings can disentangle these concepts.

Student Feedback – K-12 Education

In contrast to higher education, the literature on student feedback in k-12 settings is less extensive. Though evidence stems from only four main investigations, the promising results suggest surveys have the potential to serve as an alternative measure of teacher effectiveness. First, in a study of 1976 K-12 students in Wyoming, Wilkerson, Manatt, Rogers, & Maughan (2000) found that student ratings were significant predictors of student achievement in reading ($p<.001$) while self-ratings by teachers, principal ratings, and principal summative evaluations were not significant at even the .05 level in reading. In math, student feedback, teacher self-ratings, and principal summative evaluation were all significant. Student ratings were the only significant predictors of achievement in both Language Arts and math. The correlations from this study are shown in Table 1.

Table 1 - Correlations Among Various Measures and Student Achievement

	Math Student Achievement	Reading Student Achievement
Student Feedback	.67	.75
Teacher Self-Ratings	.67	.21
Principal Ratings	.17	.09
Principal Summative Evaluation	.51	.34

Similarly, in a study of 9,765 student surveys, researchers found that student surveys at various levels (elementary, middle, and high school) were valid and reliable teacher evaluation

measures (Peterson, Wahlquist, & Bone, 2000). This aligns with international research from Cyprus where student surveys of teacher practices were highly correlated with achievement gains in math and Greek language as well as other affective outcomes of schooling (Kyriakides, 2005). These findings "provide convincing evidence that student ratings of teaching are worth considering for inclusion in teacher evaluation systems" (Goe, Bell, & Little, Approaches to Evaluating Teacher Effectiveness: A Research Synthesis, 2008, p. 40).

More recently, an extensive research project funded by the Gates Foundation is investigating the relationship between several measures of teaching and value-added estimates of student achievement. Referred to as the Measures of Effective Teacher (MET) Project, the goal of the study was to determine the ideal components of teacher evaluation. The measures included in the study were prior value-added scores, observational rubrics, tests of teaching strategies, and student perceptions (Kane & Cantrell, Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project, 2010). The MET project operated in seven districts across the county and included more than 3000 teachers in grades 4-8.

Student perceptions are measured using the 36 question Tripod student perception survey developed by Ron Ferguson at Harvard University. The survey contains items that assess the degree to which students view the classroom environment as "engaging, demanding, and supporting their intellectual growth" (Kane & Cantrell, Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project, 2010, p. 7). It employs a 5-point Likert scale ranging from strongly agree to strongly disagree.

Preliminary findings from the project report a significant correlation between a teacher's total score on the student survey and value-added achievement on state tests in both math and ELA. These are similar to correlations between value-added and observation rubrics that look at general teaching practices such as Danielson's Framework for Teachers and

CLASS. The table below displays the correlations for each measure, with student surveys showing a .218 correlations with value-added in math and a .095 correlation with value-added in ELA. Prior to attenuation, the correlations are .093 and .057 respectively. While correlations for value-added are small and positive, it is important to note the greater reliability for student surveys. Reliability consists of the correlation between different sections taught by the same teacher, with high correlations suggesting that the measure captures characteristics of the teacher rather than idiosyncrasies within each classroom.

Table 2 - Correlations Among Measures of Teacher Evaluation and Value Added - MET Project

Measure	Correlation with Math Value-Added	Correlation with ELA Value-Added	Reliability
Student Perception Surveys – Tripod Survey (Disattenuated[1])	0.218	0.095	0.65
Student Perception Surveys – Tripod Survey (Actual)	0.093	0.057	
Danielson's Framework for Teaching	0.19	0.11	0.40
CLASS	0.24	0.10	0.43
UTOP	0.26	X	0.42
MQI	0.16	X	0.20
PLATO	X	0.20	0.38

While there does not appear to be a large difference among measures in the ability to predict value-added, there is evidence that student surveys can provide additional information above and beyond what is provided by observation rubrics. When student surveys are included, the difference in achievement gain between the top and bottom quartile teacher increases from 2.6 months of learning to 4.8 months of learning (Kane & Staiger, 2012). This suggests that a

[1] Disattenuation calculated by dividing correlation by the reliability of both value-added and the Tripod survey in an effort to correct for attenuation bias due to measurement error.

comprehensive model of teacher evaluation would be enhanced by including multiple sources of information to provide a more discriminating measure of teacher effectiveness, and that student surveys are potentially a valuable component.

Though the findings from the MET project suggest that student surveys are potentially a promising component of a teacher's evaluation, the process for development and validation of the Tripod student survey does not follow a comprehensive validation framework. There is no documentation from either MET project findings or published works about the Tripod survey that outline how it was created or the validation process. Therefore we do not know if the behaviors are related to established teaching practices or whether items have undergone cognitive testing to ensure alignment of items with question objectives. Further, at the time of this writing there is no instrument for student surveys in k-12 schools that has been created and tested following an established framework for validation. Given the potential implications of evaluating teacher performance on student surveys, having a sound theoretical support for the technical characteristics of the survey is essential.

States and Districts Incorporating Student Surveys in Teacher Evaluation

There are three states that are considering student surveys as a measure of teacher evaluation as well as at least two districts that currently use student surveys (Burniske & Meibaum, 2011). On the state level, the current investigation serves as Georgia's pilot program of student surveys within their Teacher Keys Evaluation System. Further, the Massachusetts Department of Education is in the process of selecting instruments for obtaining student feedback, with the state having surveys as one option for teacher evaluation beginning in the 2013-2014 school year (Burniske & Meibaum, 2011). Finally, the state of Arizona has recently

put out an RFP for student surveys to be used as part of a statewide component of teacher evaluation to be piloted in the 2012-2013 school year (Arizona Department of Education, 2012).

On the district level, Davis School District in Utah allows teachers to choose student surveys as one source of data for assessing teacher effectiveness, with the survey developed in 1995. This past year, Memphis City Schools adopted student surveys as a component within the district's teacher evaluation system. Although surveys represent only 5% of a teacher's evaluation, it is the first district to use this type of assessment in a high stakes setting. It should be noted that the use of student surveys is growing rapidly at the district level, with these districts representing agencies that have drawn more attention nationally.

Contributions to the Literature

The current investigation expands the existing literature on student surveys in several areas. First, it outlines the development and testing of a student survey following an established framework for validation. Previous work in student surveys has not either gone through this process or produced any documentation of evidence for construct validity.

The current study also investigates different populations and subjects in addition to looking at other outcomes. First, it extends the findings on student surveys to the high school level. Previous investigations that linked student surveys to value-added have focused mostly on middle school students. Second, it explores the relationship between student surveys and value-added in social studies and science as well as ELA and math. While ELA and math are subjects where student achievement is consistently available, it is unwise to assume that a similar relationship between teacher value-added and student surveys applies to all subjects. Next, it relates student surveys to external measures such as student engagement and self-efficacy. This

incorporates independent measures of important outcomes in education. Finally, it investigates how teachers incorporate feedback from student survey reports into their own teaching. Though valid measures of teacher effectiveness are essential; developing instruments that can both discriminate among teachers as well as make teachers more effective is an important goal and this allows for a better understanding of how teachers use the feedback in improving practice.

Chapter 3:

Methods

There are multiple issues to consider when designing a valid instrument. The first is how one should define validity; particularly since there have been varying viewpoints on the definition in the past sixty years. Two important publications have followed the developments in validity theory in education measurement. These include the *Standards for Educational and Psychological Testing* (1966) and the validity chapter in *Educational Measurement* (Moss, 2007). The 1966 publication of the Standards details three main types of validity including content validity, criterion validity, and construct validity. Content validity demonstrates how well a measure "samples the class of situation or subject matter about which conclusions are to be drawn", criterion validity compares scores with "one or more external variables considered to provide a direct measure of the characteristics or behavior in question", and construct validity seeks to determine "the degree to which the individual possesses some hypothetical trait or quality that cannot be observed directly" (APA, 1966, p. 12-13; as cited by Moss, 2007).

More recently, the 1985 Standards as well as Messick's (1989) chapter in *Educational Measurement* have presented a more unified version of validity that centers around establishing construct validity. Other forms of validity (such as content validity or criterion validity) represent evidence that supports construct validity. This belief agrees with later works (Kane, 2006) that describe true test validity as an impossible task. Instead, one needs to establish a body of evidence that support the measure's use. The following presents several pieces of evidence regarding the validation of the current student survey.

The following validation framework guides both the creation and testing of survey items in an effort to provide evidence for validity in three main areas shown in the table below.

Table 3 - Validity Framework

	Content Validity	Convergent Validity	Predictive Validity
Primary Questions to Answer	* Does it have adequate coverage?	* Is there a concurrent relationship with similar measures?	* Can survey predict similar measures?
Strategy	* Literature search on teaching practices	*Correlation with measures of academic engagement and academic self-efficacy	*Correlation with teacher value-added

Content Validity

The first question asks whether the survey has adequate coverage of effective teaching practices in an effort to establish content validity. Messick (1989) echoes earlier definitions of content validity in that he purports that it is "founded on relevance between the content of the survey and the representativeness with which it covers the domain" (Messick, 1989 as cited by Porter, Polikoff, Goldring, Murphy, Elliott, & May, 2010, p. 142). The current investigation draws upon content validity as evidence for overall construct validity. This step is required to ensure that survey items are exhaustive of potential teacher practices and reflect the most current knowledge of effective teaching.

To achieve the goal of finding what practices should be targeted, the researcher used both reviews of the literature and commonalities among established observational rubrics. A thorough literature search on effective teaching practices was conducted using Google Scholar and ProQuest using keywords such as "teacher effectiveness", "teaching behaviors", "effective

teaching practices", "Reviews of teaching practices", "effective teaching strategies", "research-based teaching", and "effective instruction". Further, studies referenced within these references were obtained as additional sources. The next step was to develop a taxonomy of teacher practices found within the reviews and code references to various teaching practices. An example of the coding procedure used is shown in Appendix B.

The current student feedback survey was developed using commonalities among established observational rubrics such as Danielson's (1996) Framework for Teaching and a thorough literature review of teacher behaviors that are found to consistently predict student achievement (Marzano, Pickering, & Pollock, 2001; Brophy & Good, 1986; Pianta, Paro, & Hamre, 2006; Schacter & Thum, 2004; Emmer & Evertson, 1981). The overall categories include presentation style, lesson structure, behavior management, productivity, teacher-student relationships, awareness of student need, feedback, challenge, engaging and motivating students, as well as content expertise.

The first procedure consisted of identifying overlapping teacher behaviors from the various reviews of the literature. For instance, all of the reviews highlight a link between providing feedback for students and higher student achievement. Schachter and Thum (2004) note that teachers should provide "frequent, elaborate, and high quality academic feedback", Good and Brophy (1986) discuss "monitoring of students' understanding and providing appropriate feedback", Emmer and Evertson (1994) note that "all student work, including seatwork, homework, and papers, is corrected, errors are discussed, and feedback is promptly provided", and finally Marzano (2001) outlines several research based feedback strategies.

When a commonality among the reviews is found, the teacher behavior is then written into a question that allows students to rate the frequency of this behavior. Table 4 displays some of the behaviors identified by the rubric and the corresponding survey questions.

Table 4 - Rubric Behavior and Corresponding Survey Question

Research Based Teaching Practice	Corresponding Student Survey Question
Feedback makes students explicitly aware of performance criteria in the form of rubrics or criterion charts.	My teacher gives us guidelines for assignments (rubrics, charts, grading rules, etc.) so we know how we will be graded.
Teacher engages students in giving specific and high quality feedback to one another.	I have opportunities during this class to give and receive feedback from other students.
The teacher circulates to prompt student thinking, assess each student's progress, and provide individual feedback.	My teacher walks around the room to check on students when we are doing individual work in class

The second procedure involved using common observational rubrics such as Charlotte Danielson's (1996) *Framework for Teaching* and the Classroom Assessment Scoring System (CLASS) for grades K-5 (Pianta, La Paro, & Hamre, 2006). Both of these instruments have been tested for validity by assessing the relationship between teacher scores on the rubric and a teacher's value-added student achievement (Kane, Taylor, Tyler, & Wooten, 2010). These also represent the two rubrics chosen to measure general teaching practice in seven large school districts as part of the current Measures of Effective Teaching project sponsored by the Gates Foundation. As such, they have been identified as valuable tools for identifying effective teacher practices. Teacher behaviors identified by the highest levels of these rubric were transformed into questions appropriate for students to answer. There was considerable overlap between the two rubrics, but certain areas were only addressed by one or the other. Examples are provided in Table 5 and the full mapping of items to rubrics can be found in Appendix C and D.

Table 5 - CLASS and Framework for Teaching Behaviors and Corresponding Student Survey Questions

CLASS	Framework for Teaching	Student Survey Question
Rules and behavior expectations are clearly stated or understood by everyone in the class.	Standards of conduct are clear.	My teacher explains how we are supposed to behave in class. I understand the rules for behavior in this class.
The teacher can answer all levels of student questions.	N/A	My teacher is able to answer students' questions about the subject.
N/A	Teacher's oral and written communication is clear and expressive and anticipates possible student misconceptions.	When explaining new skills or ideas in class, my teacher tells us about common mistakes that students often make.

Ideally, it would have possible to draw upon existing student surveys for other possible items. Unfortunately, previous student surveys do not have evidence or documentation demonstrating the link between the items and research-based teacher practices. Further, the Tripod student survey items were not available to the public at the time of development of the current survey.

The selection process was also guided by filters that asked whether students were the best judge of this behavior as well as whether students were capable of answering the question. Although the literature might suggest certain practices as components of effective teaching, the behavior must be something that students are familiar with. For instance, students may not be able to answer a question such as "my teacher plans a good lesson", but they are a good judge for questions such as "we are learning or working during the entire class period". Further, students must be able to observe the behavior. As an example, much of the literature suggests a

connection between differentiating lessons and student achievement. While this may be an important practice, students may never know that teachers differentiate their lessons and therefore these behaviors are challenging to include. Instead, it is more useful to ask about easily observable, low inference behaviors so that students can be as successful as possible.

Overall, these procedures led to the development of 64 survey questions that all have a basis in either overlapping areas of literature reviews or are grounded in descriptions of teacher behaviors from valid observational rubrics. This process provides evidence for content validity. The next step involves testing items in order to provide other sources of evidence.

Cognitive Interviews

The next source of evidence for construct validity comes from cognitive interviews. These determine whether the objectives that were noted above match how the students interpret the actual survey items. Cognitive interviews were conducted to ensure that students interpret each item according to the desired objective set forth by the researcher. These types of interviews are helpful in addressing common threats to validity associated with surveys (Porter et al., 2010; Desimone & Le Floch, 2004). Threats to survey validity arise due to complex phenomena being asking about, respondents answering in socially desirable ways, or respondents not being clear about what questions are asking; with cognitive interviews guarding against these threats. In order to respond accurately, respondents must be able to "comprehend an item, retrieve relevant information, make a judgment based upon the recall of knowledge, and map the answer onto the reporting system" (Desimone & Le Floch, 2004, p. 6). Cognitive interviews allow the researcher to determine which part of the process respondents may be having difficulty with and why.

There are two main types of cognitive interviewing (Beatty & Willis, 2007). The first, referred to as a 'think-aloud', allows respondents to verbalize the mental process as they read and answer each question. The second style takes a more active approach on the part of the researcher in which respondents are asked specific questions about survey items. The current investigation draws upon both interview types as they each offer certain advantages. Respondents used the think-aloud style as they first encountered each question in order to assess overall question clarity. There were also specific instructions to describe what teacher behaviors or experiences they are drawing upon when answering the question. If students draw on unintended teacher behaviors, follow-up questions will be asked about why the student chose these behaviors. There were also specific questions about items that are identified by the researcher as potentially confusing or ask about constructs that were challenging to translate into survey questions.

Finally, in an effort in minimize subject bias for survey items, students were asked to answer questions about teachers in a variety of different academic subjects. For instance, the first student was asked to answer questions about their math teacher, the next about their science teacher, and the next about their art teacher. Questions that did not apply to certain subjects were revised or eliminated.

In the first round, 10 students were interviewed at a local private high school in Nashville, TN. Instructions and questions that were confusing or questions that were interpreted in ways that did not address the teaching objective were revised on an iterative basis. All revisions were then tested again with a 15 student focus group at a public high school in Atlanta, Georgia. These two rounds represent an exploratory analysis focused on exposing a full range of possible problems (Blair & Brick, 2009).

There were several adjustments made as a result of these interviews. First, the original response scale included the following options: Never, Sometime, Often, Almost Always, and Every Time. As a result of repeated confusion over answering questions where "Every Time" did not apply, this option was changed to "Always". Further, some questions were eliminated based on interview feedback. Originally, there was an item that asked about dividing responsibilities while working in groups that stated "When working in groups, my teacher has us choose a job, role, or responsibility within the group (recorder, materials person, manager, etc.)". Many students felt that this question did not apply to subjects outside of science. Since the issue was with the subject of the question as opposed to the wording, the item was eliminated. Finally, other items were revised based on the results of cognitive interviews. For example, one of the items originally had the wording "When we learn something new, my teacher goes through a few examples with the class together". Several students noted that "a few" was confusing so the item was reworded to state "When we learn something new, my teacher goes through examples with the class together".

Further interviews were conducted with both former teachers and content experts. First, five former teachers were interviewed and asked to read the question, describe whether the question was understood, state what objective the question is likely trying to address, and finally, provide an assessment of how well the question addressed that objective. Following these interviews, several questions were revised, elaborated, or eliminated based on clarity and ability to match survey question with intended objective. Additionally, four content experts were provided with the survey and asked to provide feedback on whether the questions covered an adequate range of teacher behaviors, whether the questions were asking about important teacher behaviors, and how questions might be improved. Again, questions were revised based on this feedback.

Response Scale

An important characteristic of any survey is the response scale. In an effort to make the questions as objective as possible, the scale was designed to have students rate the frequency of low-inference behaviors. Murray (1983) investigated the questions that asked about both high-inference and low-inference teacher behaviors on student surveys. High-inference questions such as "Is the instructor clear?" or "Does the teacher plan a good lesson?" are not able to communicate information about actual teacher behaviors in a classroom. On the contrary, questions regarding low-inference behaviors require less judgment on the part of the observer, thus allowing students to rate them more objectively. Instead of asking about instructor clarity, a related question concerning a low-inference behavior might ask the student to rate the frequency of whether "My teacher uses examples or illustrations to help explain ideas". By asking questions about concrete behaviors that are easy to identify in addition to asking about the frequency of behavior, the validity and reliability of student surveys improves. The survey therefore uses a rating scale from 1 to 5 that asks about the frequency of teacher behaviors. The rating scale categories include 'Always', 'Almost Always', 'Often', 'Sometimes', and 'Never'.

Scale Development

Scales for the survey are connected to previous constructs within the field of teacher effectiveness. While previous student surveys have not had scales, some guidance comes from the structure of observation rubrics. The table below outlines the relationship between previous scales from Schachter and Thum (2004) as well as from the CLASS rubric (Pianta, Paro, & Hamre, 2006). These rubrics are particularly good examples of scales of rubrics that organize teacher behaviors into large categories rather than having one overall scale for teacher effectiveness.

Table 6 - Construct Alignment with Observational Rubrics

Previous Constructs		Current Student Survey Constructs	
Schachter and Thum Constructs	**CLASS Contruct**	**Teacher Role Sub-Category**	**Teacher Roles**
Presentation	Instructional	Presentation Style	Presenter
Lesson Structure and Pacing, Lesson Objectives	Learning Formats	Lesson Structure	
Classroom Environment	Behavior Management	Behavior Management	Manager
	Productivity	Productivity	
Classroom Environment	Positive/Negative Climate	Teacher-Student Relations	Counselor
Teacher Knowledge of Students	Teacher Sensitivity	Awareness of Student Need	
Feedback	Quality of Feedback	Providing Feedback	Coach
Activities	N/A	Challenging Students	
Motivating Students	Regard for Adolescent	Investing Students	Motivator
Questions	Perspective	Engaging Students	
Teacher Content Knowledge	Content Understanding	Content Knowledge	Content Expert
Thinking	Analysis and Problem Solving	Encouraging Thinking	

Each of the sub-categories for the survey scales has a connection to previous scales in the two rubrics. These sub-categories were then combined into larger categories that were developed by the researcher. Since one of the goals for the research project was to provide feedback to teachers from student surveys, the categories were grouped in a way that was meaningful to teachers. Teachers can relate to the fact that they are asked to play many roles as a teacher, often within the same class period. Having the feedback organized in a way that is intuitive for teachers could potentially allow for a better reception and comprehension of the feedback. As these larger categories are previously untested, the investigation will include a confirmatory as well as an exploratory factor analysis to determine optimal alignment.

Pilot Testing

Sample

Pilot testing took place in the spring of 2011, with the majority of work conducted in Georgia as part of the state's Race to the Top initiative. With assistance from the researcher, Georgia included student surveys as a component of a teacher's evaluation in their Race to the Top application. The office charged with writing the application was the Governor's Office of Student Achievement. Upon being awarded the grant, the state agreed to participate in a large-scale pilot study to validate the current student survey.

The sampling frame includes seven districts that represent urban, suburban, and rural districts, with basic information on each district provided below. The choice of districts was dictated by Race to the Top staff and district level sampling . Georgia's Race to the Top office had previously divided the 26 participating districts into three separate groups for instrument testing in the areas of value-added, observation rubrics, and student surveys. Although each group of districts was divided to include a diversity of district characteristics, the results from the study can only generalize to the seven participating districts.

Table 7 – 2008-2009 Demographic Information for Districts in Georgia

	Number of Students	Percent Eligible for Free/Reduced Lunch	Percent Limited English Proficiency	Urbanicity	Number of Schools
DeKalb	99,775	66.1%	7.5%	Suburb: Large	48
Griffin-Spalding	10,823	66.7%	1.0%	Suburb: Large	10
Hall	25,629	53.5%	17.5%	Rural: Fringe	12
Meriwether	3,449	81.0%	0.8%	Rural: Distant	4
Pulaski	1,593	60.1%	1.3%	Town: Distant	2
Rabun	2,244	60.7%	5.8%	Rural: Remote	2
Savannah-Chatham	33,994	61.8%	1.8%	City: Mid-size	17

All middle and high schools within each of the districts participated, but selection strategy of teachers varied by district. For smaller districts, all teachers within the districts participated in the pilot. In larger districts, all schools participated in the pilot but teachers were randomly sampled (RS) within schools based on availability of teachers and capacity of the district. The strategy and resulting number of teachers is shown in

Table 8 below.

Table 8 - Sampling Strategy and Resulting Number of Teachers

	Schools	Sampling Strategy (Teachers Per School)	Teacher Response Rate	High School Students	Middle School Students
DeKalb	48	5 RS	121/240 (50.4%)	1,156	1,215
Griffin-Spalding	10	20 RS (HS) 10 RS (MS)	75/160 (47%)	712	625
Hall	12	15 RS	166/180 (92%)	1,674	1,663
Meriwether	4	All	65/89 (73%)	728	555
Pulaski	2	All	39/50 (78%)	390	367
Rabun	2	All	68/87 (78%)	634	369
Savannah-Chatham	17	10 RS	133/163 (82%)	1,265	1,055
Total	95		667/889 (75%)	6,559	5,849
		Total Students		12,408	

In two of the districts (DeKalb and Griffin-Spalding), there were technical difficulties with the online administration that resulted in lower response rates. Specifically, in one case the district was late in removing a bandwidth filter that led to several teachers having students who

could not access the website. In another situation, over 4000 students took the survey on one day. Since this was larger than the anticipated need for server space, some students could not access the website in the time required to switch to an unlimited capacity server. It is unknown how these factors affected the sample of teachers, but if certain types of teachers were prevented from having their students access the website then the results would be biased. Considering these issues are outside factors likely unrelated to teacher effectiveness, it is also possible that data are missing at random. Still, this should be considered a limitation of the current study.

Measures

Academic Engagement

Student engagement examines student's report on their interest in learning. The measures for the current investigation were developed and tested by the Consortium on Chicago School Research (CCSR) with more than 100,000 demographically diverse elementary and high school students in Chicago Public Schools (Fredricks & McColskey, 2011). The 4-point Likert scale ranges from 'Strongly Agree' to 'Strongly Disagree' and includes six questions. Overall summary statistics for high school include individual separation (1.37), individual level reliability (.65) and school level reliability (.88). Item characteristics of are provided below.

Table 9 – CCSR Measure of Academic Engagement

	Item Difficulty	Item Fit
The topics we are studying are interesting and challenging	0.54	0.71
I am usually bored in this class	0.76	0.89
I usually look forward to coming to this class	0.76	0.57
I work hard to do my best in this class	-0.37	0.88

Sometimes I get so interested in my work I don't want to stop	0.93	0.75
I often count the minutes until class ends	1.18	1.07

Academic Efficacy

Academic efficacy refers to student perceptions of their competence to do their class work. It was developed as part of the Patterns for Adaptive Learning Scales (PALS) survey at the University of Michigan. The scales are based on research showing that an emphasis on mastery rather than performance is related to more adaptive patterns of learning (Midgley et al., 2000). Items were tested in nine districts in three Midwestern states at the elementary, middle, and high school level. The five question scale uses a 5-point Likert rating, and has a Cronbach alpha score of .78.

Table 10 – PALS Measure of Academic Self-Efficacy

	Mean	**SD**
I'm certain I can master the skills taught in class this year.	4.17	0.94
I'm certain I can figure out how to do the most difficult class work.	4.10	1.04
I can do almost all the work in class if I don't give up.	4.42	0.92
Even if the work is hard, I can learn it.	4.42	0.90
I can do even the hardest work in this class if I try.	4.33	1.04

Teacher Value-Added

The relationship of student surveys to estimates of a teacher's value-added scores will help provide evidence for criterion validity as gains in student achievement are arguably the most common metric for performance in education. Given the alignment of behaviors on the survey to those that have previously demonstrated a relationship to student achievement, one would expect that a greater frequency of these behaviors would be associated with larger gains in achievement in the current study.

To calculate value-added scores for teachers, a model adapted from the MET project will be employed (Kane & Staiger, 2012). Model 1.1 includes the achievement of student *i* of teacher *k* as the outcome, a student's prior achievement, a grade fixed effect, and student characteristics that may influence achievement (examples include free and reduced price lunch status and race). The error terms represent unexplained variance at the student level (ε).

$$(1.1) \quad A_{ijk} = \beta_o + \beta_1 A_{ijk\,t-1} + \beta_2 X_{ijk} + \beta_3 X_{jk} + \eta_j + \varepsilon_{ijk}$$

$A_{ijk\,t-1}$: Student Prior Achievement
X_{ijk}: Race, FRL Status, ESL Status, Special Ed Status
X_{jk}: Classroom Means for Demographics
η: Grade Fixed-Effect

Factor Analysis

Factor analysis looks for systematic relationships among multivariate data in order to "identify a limited number of interpretable, unobserved variables that explain the meaningful covariation among a set of observed variables" (Preacher, 2012, p. 6). Factor analysis can either provide evidence for existing scales in a confirmatory factor analysis or explore data for possible relationships in an exploratory factor analysis. Since there is a strong connection between survey constructs and previously validated scales, a confirmatory factor analysis is appropriate as this allows the researcher to test pre-specified groupings of items. Still, an exploratory analysis can identify alternative grouping structures that could improve the functionality of the survey. Both analyses are conducted in the current study.

Item Reliability/Discrimination

Item discrimination provides additional evidence of survey reliability by measuring the relationship between individual items and a teacher's total score. Items that have either no relationship or a negative relationship may undermine validity as the item may be measuring

something other than intended. Item discrimination will be calculated using a Spearman correlation between item score and a teacher's total score. This test is preferable to using Pearson's correlation because it is unknown whether the relationship between each question and the total score should be expected to be linear.

Possible Threats to Validity

There are potential factors that may detract from the survey validation. First, it is possible that students may not spend adequate time answering survey questions. This could result in students putting random answers that may have no relationship to the actual frequency of teacher behavior. To prevent this, answers that fall 1.5 standard deviations away from the class mean will be flagged. Though this discrepancy may have meaning at the individual question level (for instance, if a teacher did not check for understanding with all students), a repeated pattern of deviance from the class mean may indicate that the student was not taking the survey seriously. Therefore, students who have more than 1/3 of their answers flagged will be checked for repeated, consecutive answers or suspicious answer patterns.

Next, a possible threat to validity is the time that a child spends in a teacher's classroom. A student may have a biased opinion of a teacher if they have not had adequate time to observe the variety of behaviors that are asked about in the survey. While there is no specified minimum number of days that a student needs to attend to observe a full range of a teacher's behaviors, it is reasonable to assume that a student has had enough time to observe the teacher if they have spent more than a month in their classroom as the behaviors listed on the survey should be observed on a regular basis. The survey will therefore include a preliminary question that asks the students how long they have been enrolled in this teacher's class. Students that answer 'less than 1 month' will be excluded when calculating a teacher's total score.

A further threat would be that student characteristics may influence ratings. For instance, there is some evidence that students rate female teachers higher (Aleamoni, 1999). To check for this, student level controls for gender, race, ethnicity, and socioeconomic status will be investigated for their influence on student ratings.

Finally, it is possible that teachers may try to influence their ratings based on which students take the survey. Part of the research design reduces this likelihood since the class that is surveyed was randomly chosen from all of a teacher's classes. Still, teachers may try to manipulate which students within the sampled class actually take the survey. To minimize the incentives for the type of behavior, teachers were consistently told that individual results would not be shared with school, district, or Race to the Top administrators. This message was relayed in messages from district staff, a survey introduction letter to all teachers, as well as the actual survey instructions. Unfortunately, it is not possible to tell which students did not participate in the survey since it was anonymous at the student level.

Chapter 4:

Results

Data

Survey data were collected in the spring of 2011. Overall, 12,944 students complete the survey. Some of the online surveys were incomplete due to technical issues both at the district level as well as a temporary issue with server capacity. Of these surveys, 12,408 (95.9%) were able to be matched with teachers in the sample. Table 11 displays the number of students taking surveys within each of the seven districts.

Table 11 – Student Sample by District

District	Number of Students Completing Survey	% of Total Sample
DeKalb	2,361	19.03
Griffin-Spalding	1,337	10.78
Hall	3,399	27.39
Meriwether	1,229	9.90
Pulaski	757	6.10
Rabun	1,003	8.08
Savannah-Chatham	2,322	18.71
Total	12,408	100.0

Of students taking the survey, 47.1% were in middle school with the remaining 52.9% enrolled in high school. A further breakdown of students by grade is displayed in Table 12.

Table 12 – Student Sample by Grade

Grade	Number of Students	% of Students
6	1,698	13.70
7	2,257	18.21
8	1,882	15.18
9	1,824	14.71
10	2,011	16.22
11	1,643	13.25
12	1,080	8.71
Total	12,408	

In terms of race/ethnicity, the composition of the sample was predominantly African-American or White/Caucasian. Table 13 shows the number and percent of students within each race.

Table 13 – Student Sample by Race/Ethnicity

Race/Ethnicity	Number of Students	% of Students
African American	4,581	36.9
Asian	345	2.8
Hispanic	1,614	13.0
White/Caucasian	4,965	40.0
Other	903	7.3
Total	12,408	

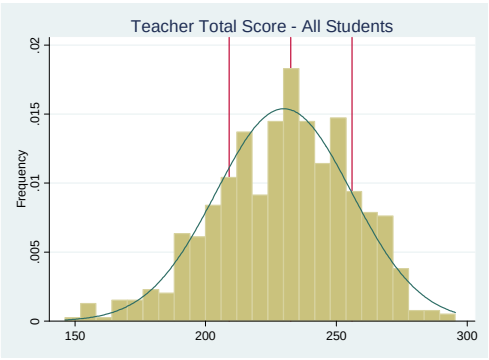
Distribution of Scores

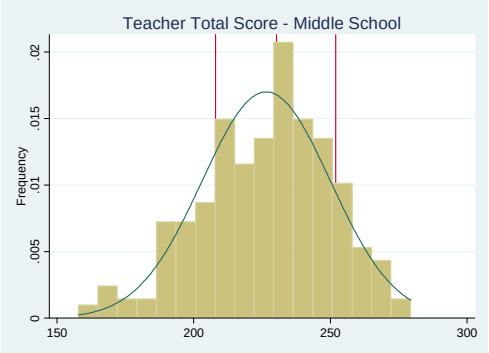
Totals from the 64 questions were added together to produce a teacher's total score. This score could range from a minimum of 64 to a maximum of 320 and overall averages for each level are shown in Table 14.

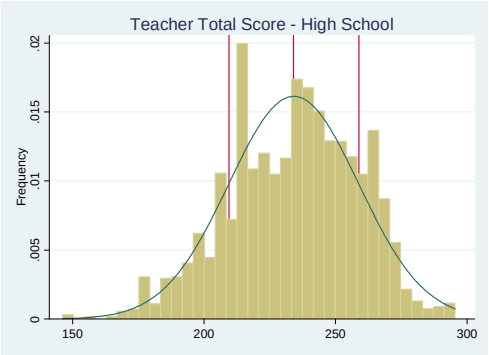
Table 14 - Descriptive Statistics for Teacher Total Score and Teacher Total Average

	Overall	Middle School	High School
Teacher Total Score	229.92	226.56	231.76
	(25.92)	(23.46)	(29.51)
Teacher Total Average	3.63	3.60	3.66
(1-5)	(0.37)	(0.34)	(0.39)
Number of Students	12,408	5,841	6,567
Number of Teachers	667	294	373

The figures below show the distribution of teacher total scores for all teachers, those in high school, and teachers in middle school. These distributions are approximately normal. The middle red line represents the mean, with lines on either side displaying one standard deviation above and below the mean.







Screening Procedures

Before a total score is calculated for a teacher, statistical techniques can assist with identifying surveys that may be invalid. For instance, teachers expressed concern over the potential situation where students do not take the survey seriously and either choose the same answer for every question or intentionally answer questions in a way that alters a teacher's overall score. To allay these fears, screening procedures were used to identify and eliminate invalid survey responses.

The first screening procedure identifies answers that have a minimum difference in comparison to a teacher's total average for that particular question. The standard deviations for items range from 1.10 to 1.35. As such, a difference of only one standard deviation could potentially flag answers that are still quite close to a teacher's average. An example would be a teacher's average being 2.85 and the student choosing a 4 resulting in a flagged answer. Traditionally, two standard deviations represent data that falls within the 95% confidence interval. However, using two standard deviations would prevent questions having an average between 2.3 and 3.7 from ever having either extremely high or low answers flagged. As such, 1.5 standard deviations was chosen as the minimum difference needed for an answer to be flagged. The resulting figure results in a confidence interval of 86.6%. This is roughly the minimum difference that allows questions of any average to still have flagged answers as the maximum value is close to 2 (1.35*1.5 = 2.025). For example, a question with the average of 3.0 and a standard deviation of 1.32 will still be flagged if a 5 or a 1 is chosen.

If a survey has a minimum number of 25 flagged answers then the survey was eliminated. Upon closer investigation, the surveys that had at least 25 flagged answers generally contained repeated answer strings. Table 15 shows the number of surveys that were eliminated using this procedure and the percentage of students that had each range of flags.

Table 15 - Number of Eliminated Surveys

	Number of Students	% of Students
Valid Surveys	11,786	94.99
Eliminated Surveys	622	5.01
Total	12,408	

Table 16 - Number of Flags

Number of Flags	Percent of Students
0-10	84.95%
11-15	5.65%
16-20	2.77%
21-25	2.03%
26-30	1.26%
31-35	1.04%
36-40	0.59%
41-45	0.61%
46-50	0.49%
51-55	0.39%
56-60	0.04%

A further screening mechanism involved student response to the question "I was being honest when taking this survey". Surveys with answers of either Disagree or Strongly Disagree were eliminated. This technique has been previously found to identify false responses in student surveys (Reniscow et al., 2001; GAO, 1993). The table below displays responses for this screening question, with roughly 6% of surveys being eliminated based on this question.

Table 17 - Student Responses to "I was being honest when taking this survey"

	Number of Students	% of Students
Strongly Agree	7,361	73.5
Agree	2,203	22.0
Disagree	361	3.6
Strongly Disagree	254	2.5

While a high degree of crossover occurred between the two screening procedures, 1.6% of surveys were identified by only the honestly question and 3.8% of surveys were identified only by the class average screen (see table below). This suggests that each technique was identifying a unique characteristic of survey takers. Other screening procedures to consider in the future would be minimum time spent on the survey (in an online administration) as well as scores for negatively and positively worded answers about the same teacher practice.

Table 18 - Number of Students Identified by Screening Procedures

	Number of Students	% of Students
Total Students	12,408	100%
Identified by Either Class Average or Honesty Question	1093	8.8%
Identified by Honesty Question Only	200	1.6%
Identified by Class Average Screen Only	471	3.8%
Identified by Both	422	3.4%

Overall, the number of teachers stays consistent before and after screening procedures are applied as there was no case where all of a teacher's surveys were eliminated. Still, the number of students for each teacher is reduced. Further, since test scores are not available for all

teachers there is a smaller sample for the analysis that investigates the relationship between a teacher's total score on the survey and value-added. The table provides information on the number of teacher and students that included within each analysis.

Table 19 - Number of Teachers and Students in Full and Reduced Sample

	Teachers	Students
Full Sample	667	12,408
Sample after Screening Procedures	667	11,515
Value-Added Sample	360	7,214
Value-Added Sample after Screening Procedures	360	6,713

Survey Properties

Internal Consistency

The existing scales will be tested for internal consistency using Cronbach alphas. Cronbach alpha measures how closely a set of items are related together as a group. Generally, alpha levels above .7 indicate that items have adequate internal consistency. Table 20 displays the Cronbach alpha scores for each scale as well as the number of items. Overall, all of the scales display the desired levels of internal consistency, suggesting that questions within each construct are measuring similar aspects of teacher quality.

Table 20 - Cronbach Alpha Values for Survey Scales

	Presenter	Manager	Counselor	Coach	Motivator	Expert
Cronbach Alpha Score	0.893	0.704	0.821	0.824	0.850	0.820
Number of Items	13	9	10	13	10	8

Part of these high alphas comes from the fact that the scales are highly correlated with each other. The table below shows the correlations among the different scales.

Table 21 – Correlations Among The Six Survey Scales

	Presenter	Manager	Counselor	Coach	Motivator	Expert
Presenter	x					
Manager	.746	x				
Counselor	.886	.767	x			
Coach	.897	.757	.884	x		
Motivator	.927	.730	.886	.889	x	
Expert	.918	.717	.815	.886	.899	x

These high correlations suggest that each scale is potentially measuring one underlying construct of overall teacher effectiveness. However, the fact that most correlations are below .9 suggest there may be important differences between the different scales. As such, factor analysis can provide further evidence about the proper survey organization.

Exploratory Factor Analysis

Exploratory factor analysis is a technique that investigates the underlying structure of the relationship between variables in a dataset. While the scales used in the current survey draw from meaningful constructs in previous research, it is potentially useful to analyze the data without prior assumptions. The following details the result of an exploratory factor analysis using an oblique rotation in Stata. Oblique rotation, in contrast to orthogonal rotation, allows for

correlation to exist among factors. This is particularly relevant to teaching practices since high quality teachers are likely to demonstrate effective teaching practices in many different areas.

The table below displays the Akaike's information criterion (AIC), Bayesian information criterion (BIC), and Eigenvalue for each number of potential factors. The lowest BIC value appears at 16 factors, with the value increasing for greater number of factors after 16. According to the Kaiser criteria, only two factors have an Eigenvalue higher than 1. This rule, however, should not be treated as an exact science. Instead, it is desirable to find a noticeable drop-off point where factors explain less of the data.

Table 22 - Factor Analysis Results to Determine Number of Factors

Number of Factors	AIC	BIC	Eigenvalue
1	41703	42095	21.303
2	34314	35090	1.191
3	28381	29535	.925
4	22902	24426	.866
5	18101	19987	.666
6	14653	16896	.515
7	12378	14970	.508
8	10551	13484	.392
9	9111	12379	.306
10	7656	11251	.239
11	6921	10837	.217
12	6344	10572	.177
13	5775	10310	.154
14	5245	10079	.148
15	4798	9923	.116
16	4493	9904	.092
17	4226	9915	.075

Importantly, it appears that there is one main underlying factor that potentially represents general teaching ability. Evidence for this comes from the large eigenvalue for a single factor (21.303) in comparison to the remaining eigenvalues for all other potential choices for the number of factors. This is relevant in the context of calculating a teacher's overall score since it

indicates that relationship between each individual item and this primary factor. However, for the purpose of giving feedback, the scales derived from prior theory have meaning to teachers and are helpful in giving context to teaching behaviors instead of listing several seemingly unrelated practices. This secondary goal of providing feedback is supported by the subscales developed by the researcher. Therefore a confirmatory factor analysis will provide insight on whether these categories could be used.

Confirmatory Factor Analysis

Each of these scales was also investigated using confirmatory factor analysis. Confirmatory factor analysis investigates the correlation of items within the same scale and tests for various indices of fit. Since scales related to constructs from prior research and theory, it is useful to maintain this structure if it fits the existing data. Confirmatory factor analysis was implemented using the 'factor' command in Stata. The table below displays the fit indices for the confirmatory factor analysis that includes all items.

Table 23 - Confirmatory Factor Analysis Results

	Presenter	Manager	Counselor	Coach	Motivator	Expert
RMSEA	.0487	.1126	.1260	.0644	.0900	.0503
CFI	.9704	.6166	.8209	.9103	.9087	.9735
Number of Items	13	9	10	13	10	8

Generally, it is preferable to have scales with an RMSEA of lower than .1 and a CFI of greater than .9. When using all items, it appears that both the Manager and Counselor scales do not meet these criteria. However, it does appear that students had trouble with negatively

worded items throughout the survey. The scales were rerun without these items (see table below), and all scales now fall within the proper range.

Table 24 - Confirmatory Factor Analysis Results with Revised Scales

	Presenter	Manager	Counselor	Coach	Motivator	Expert
RMSEA	.0483	.0599	.0941	.0646	.0995	.0503
CFI	.9758	.9631	.9408	.9250	.9127	.9735
Number of Items	12	6	8	12	9	8

While there is evidence that there is one main underlying factor, the fit indices in confirmatory factor analysis suggest that the scales originating from previous theory are relevant. Given their utility in providing feedback and connection to earlier research, it is preferable to maintain the existing structure. It is import to note, however, that the survey appears to be asking questions that all relate to one overall teacher effectiveness construct.

Relationship to Outcome Measures

Student Academic Engagement and Self-Efficacy

The three outcomes used in this investigation include two measures that were administered concurrently with the survey as well as a teacher's value-added scores. The two concurrent measures are a 6-question index of academic engagement and a 5-question index of academic self-efficacy. For the first outcome, correlations between a teacher's total score and academic engagement as well as academic self-efficacy are displayed in the table below. The total score is calculated by adding a student's total for a teacher and correlating this score with their own totals for each scale.

Table 25 - Correlations Between Survey Total and Academic Engagement and Self-Efficacy

	Engagement		Self-Efficacy	
	Full Sample	**With Screening Procedures**	**Full Sample**	**With Screening Procedures**
Overall	.7199***	.6750***	.6411***	.5712***
Middle School	.7000***	.6574***	.6528***	.5839***
High School	.7374***	.6956***	.6359***	.5650***
Number of Students	12,408	11,515	12,408	11,515

* p < .10, ** p < .05, *** p < .001

Overall we see very high correlations between a student's total score for a teacher and their level of reported academic engagement and self-efficacy. The correlations are slightly higher for engagement than academic self-efficacy. The interpretation here is that students with teachers who adopt the practices asked about by the survey have students that are more engaged in the class and report a greater level of confidence in the subject.

These numbers are potentially upward biased due to the fact measures were administered concurrently and a student could have developed a response pattern (e.g. all high responses or all low responses). Therefore, these correlations should represent the upper bound of the true number. It does appear, however, that screening procedures correct some of the upward bias. Since the screening procedures often identified students that marked the same answer for all questions, removing these surveys could be expected to lower the correlation between measures on the same survey.

The next analysis looks at the relationship between teacher scores within each of the survey scales and students' report of engagement and self-efficacy. The results are displayed in the tables below.

Table 26 - Correlations Between Survey Total, Scale Scores and Academic Engagement

Engagement	Teacher Total Score	Presenter	Manager	Counselor	Coach	Motivator	Expert
Overall	.675***	.631***	.525***	.612***	.610***	.697***	.548***
Middle School	.657***	.609***	.529***	.614***	.598***	.661***	.502***
High School	.696***	.660***	.515***	.627***	.632***	.733***	.594***

* p < .10, ** p < .05, *** p < .001

Table 27 - Correlations Between Survey Total, Scale Scores and Academic Self-Efficacy

Self-Efficacy	Teacher Total Score	Presenter	Manager	Counselor	Coach	Motivator	Expert
Overall	.571***	.555***	.419***	.502***	.521***	.550***	.497***
Middle School	.584***	.567***	.448***	.505***	.536***	.550***	.507***
High School	.565***	.560***	.388***	.500***	.516***	.559***	.504***

* p < .10, ** p < .05, *** p < .001

Each of the scale scores for teachers show a very strong relationship to both engagement and self-efficacy with all correlations being positive and significant. For engagement, the data follows an intuitive pattern, with the strongest relationship coming from a teacher's ability to motivate students and the weakest relationship between classroom management and engagement. In both cases, a teacher's ability to present information shows one of the stronger relationships with both measures.

Value-Added Estimates of Teachers' Contribution to Student Achievement

A further outcome of interest is gains in student achievement. While we would expect there to be a relationship between teacher scores on a student survey and a teacher's value-added, it is important to frame expectations for the relationship. The literature on value-added consistently finds a large amount of error in value-added calculations (MacCaffrey, J.R., Koretz, & Hamilton, 2003; Guarino, Reckase, & Wooldridge, 2011). Further, even value-added measures using the same test do not show high correlations between sections or between prior and current year scores. Findings from the MET project show a .380 correlation in math between value-added for the same teacher in different sections, and a .404 correlation with value-added from the prior year. For ELA, the correlation among different sections is .179 and the correlation with the prior year is .195 (Kane & Cantrell, 2010).

In addition, research indicates that more other measures of teacher effectiveness such as observation rubrics and student surveys show a small, positive relationship to value-added with correlations ranging between .1 and .25 (Kane & Staiger, 2012). It is also possible that this low correlation is attenuated by the low levels of reliability (See Table 2). While we would still expect there to be a positive relationship between effective teaching practices and value-added since many of the behaviors on the survey have previously demonstrated this relationship, it is likely that student surveys are also measuring something different. Therefore we would expect similar small, positive relationships between teacher total score on the current survey and value-added student achievement.

Before calculating value-added, a series of data rules were devised in order to ensure that prior test scores are predictive of current year data. One issue is that End of Course Tests administered in high school are not vertically aligned with CRCT tests administered in middle school. Therefore scores were standardized so that prior tests may be used to predict current

year scores. The following represents that breakdown of current year tests in the student achievement database. The EOCT 9LC and ALC are both in ELA and MAT1 while MAT2 are both in math.

Table 28 - Number of Students Taking Each Test

	Number of Students	% of Students
CRCT	7,893	50.57
EOCT 9LC	1,266	8.11
EOCT ALC	1,133	7.26
EOCT MAT1	2,273	14.56
EOCT MAT2	3,042	19.49

The literature on high school value-added presents a challenging picture of identifying appropriate prior test scores (Goldhaber, Goldschmidt, Sylling, & Tseng, 2011). Still, though tests are not vertically aligned it is possible to standardize scores in order to exploit the available data. Further, the analysis will look at the relationship in middle school (where CRCT is available) only, high school only, and a combined analysis. Value-added was calculated in a regression framework by predicted residuals from a model that included a student's prior test score that was standardized for the grade level in that year. It also included demographic information such as student race, free/reduced lunch status, special education status, ELL status, and gifted status as well as class level averages for these characteristics. The residuals from this model were correlated with a teacher's total score on the student survey to assess their predictive validity.

The table below displays the results for math and ELA. It shows the relationship when using the full sample, the relationship when surveys are removed using the screening procedures described above, and when negatively worded items are removed from a teacher's total score. As demonstrated earlier in the factor analysis, it appears that students had more difficulty with

negatively worded items. Questions such as "My teacher presents material too fast for me to understand well" or "When my teacher asks questions, he/she only calls on students that volunteer" both showed negative correlations with value-added as well as student engagement and efficacy. Therefore it is helpful to investigate whether there are changes when these items are removed.

Table 29 - Correlation Between Survey Total and Value-Added Scores in Math and ELA

	Math			ELA		
	Full Sample[2]	**Screening Procedures Included**	**No Negative Items**	**Full Sample**	**Screening Procedures Included**	**No Negative Items**
Overall	.1632*	.1624* (n=110)	.1657*	.1773*	.1917* (n=86)	.1780*
Middle School	.0225	.0452 (n=54)	.0280	.2204*	.2358* (n=49)	.2203
High School	.2634*	.2507* (n=56)	.2639*	.1286	.1434 (n=37)	.1306

* p < .10, ** p < .05, *** p < .001

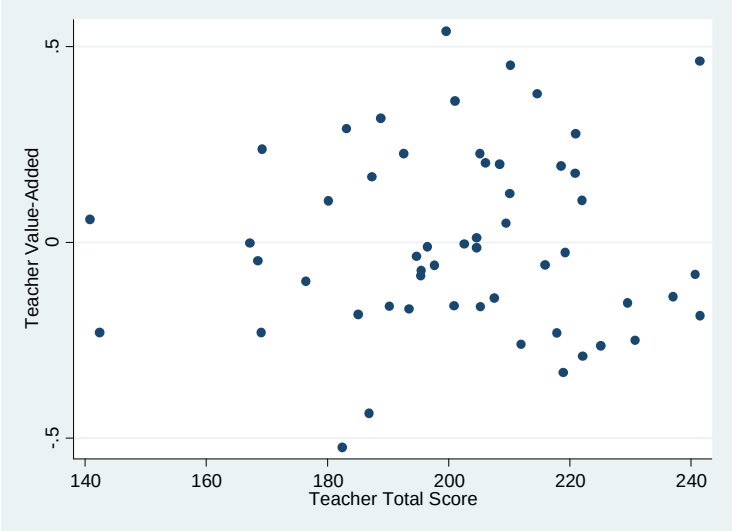
In looking at the value-added scores for teachers we also see a small positive relationship with a teacher's total score. In high school math and ELA, these correlations are significant at the p<.1 level. It math there was not a large change when either the screening procedures were implemented or with negatively worded items removed. Further, there appears to be a large difference between the correlation in math at the middle school and high school level. One possible explanation is that high school students are better able to make judgments about

[2] Full sample refers to teacher's with student achievement scores (n=360) rather than the full sample of teachers (n=667)

teachers due to better comprehension or maturity. This hypothesis, however, is not supported by evidence in ELA (or by results in science shown later).

One concern is the relationship between value-added in math and a teacher's total score for middle school since it is lower than all the other values. There are a range of possibilities that could influence this correlation that will be investigated including outliers, the distributions, regression diagnostics, and a district level analysis. In first looking at outliers, the graph below shows the plot of teacher value-added and teacher total score for middle school math. It does not appear that the data is being skewed by any larger outlier (which would be a single unit in either the top left of the bottom right quadrant). There do appear to be several teachers grouped in the bottom right quadrant that have a more negative relationship.

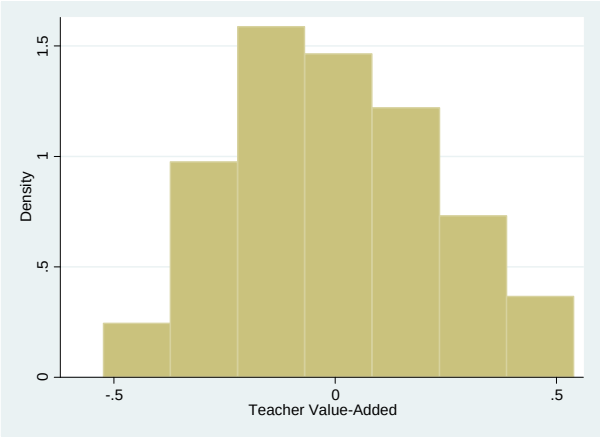
Figure 1 - Graph of Teacher Value-Added and Total Score for Middle School Math



The next possibility is having a skewed distribution for value-added within middle school math teachers. The figure below shows the distribution for teacher average value-added.

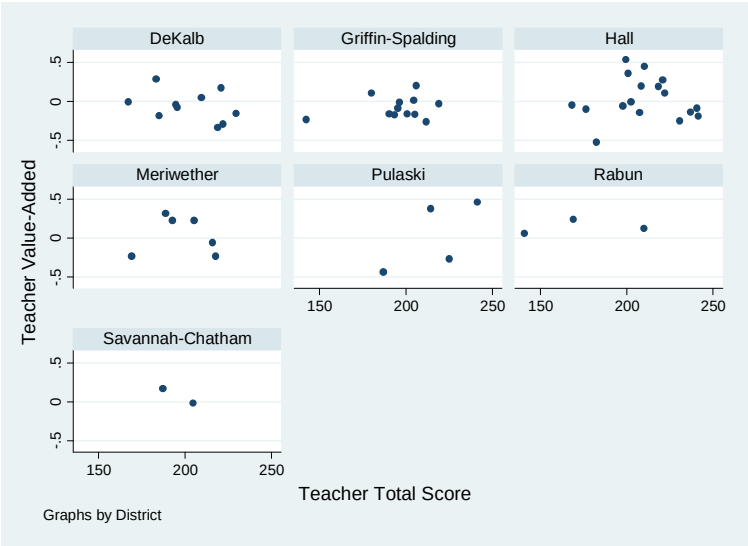
Although there is a bit of shift the left, the distribution does approximate normal and likely is not responsible for the discrepant middle school results.

Figure 2 - Distribution of Value Added in Math for Middle School Teachers



The next figure investigates each of the correlations at the district level. It does appear that one district in particular contributed to the lower correlation with math at the middle school level. The correlation for the 10 middle school math teachers in DeKalb County is -.3643, with the overall middle school correlation in math rising to .1095 when these teachers are removed. Although it is beyond the scope of this paper, several middle schools in DeKalb were indicated in a major cheating scandal in 2010 and 2011 (Georgia Governor's Office of Student Achievement, 2011) which could potentially bias correlations.

Figure 3 - Teacher Value-Added in Math and Teacher Total Score by District



The next analysis looks at the relationship between a teacher's total score on the survey and value-added student achievement in science and social studies. The table below again shows the results for the full sample, the sample after surveys have been removed by the screening procedures, and the correlations with a teacher's total score that does not include negatively worded items.

Table 30 - Correlation between Survey Total and Value-Added Scores in Science and Social Studies

	Science			Social Studies		
	Full Sample	**Screening Procedures Included**	**No Negative Items**	**Full Sample**	**Screening Procedures Included**	**No Negative Items**
Overall	.2059*	.1895 (n=72)	.2025*	.3043**	.2974** (n=75)	.3046**
Middle School	.2921	.2440 (n=26)	.2852	.2713*	.2500 (n=37)	.2658
High School	.2248	.2162 (n=46)	.2223	.3460**	.3456** (n=38)	.3497**

* p < .10, ** p < .05, *** p < .001

For science and social studies there appears to be a stronger relationship between a teacher's total score and value-added student achievement. For science this is significant at the p<.1 level while in social studies the correlation is significant at the p< .05 level. In science there is a stronger relationship at the middle school level although both are relatively similar. For social studies, there is a stronger relationship at the high school level which is significant despite the low sample size.

The next step is to look at correlations with value-added for each of the six scales. The table below shows the correlations for all subjects. Overall it is clear that a teacher's ability to present information is consistently related to greater value-added for a teacher. This scale includes questions such as "My teacher presents information in a way that makes it easy for me to understand" so it is probably not surprising that this has a relationship for each subject. The second most consistent relationship is for content expertise, with significant relationships in ELA, science, and social studies. This aligns with previous research indicating an important link

between teacher content knowledge and student achievement (Greenwald, Hedges, & Laine, 1996; Rowen, Correnti, & Miller, 2002; Ferguson, 1991). Although scales such as counselor, coach, and motivator do not show significant relationships in all subjects, it is possible that the benefits to these behaviors arises in non-academic ways similar to the strong relationship shown earlier between student academic engagement and the motivator scale.

Table 31 - Correlations with Value-Added by Survey Scale

	Teacher Total Score	Presenter	Manager	Counselor	Coach	Motivator	Expert
Math	.1624*	.1825*	.1491	.1913**	.1529	.1511	.1422
ELA	.1917*	.1841*	.1141	.1454	.1668	.1393	.1852*
Science	.1895	.2040*	.2440**	.1811	.1805	.2228*	.2143*
Social Studies	.2974**	.2988**	.3475**	.2537**	.2893**	.2718**	.3201**

* p < .10, ** p < .05, *** p < .001

Item Characteristics

One consideration in using student surveys as a measure of teacher effectiveness is whether items function in similar ways for different groups of students or students that may receive different grades. Determining whether an item functions differently for a certain group can be troublesome as it is difficult to assess whether certain types of students answer the question differently or whether certain types of students have access to varying levels of teacher quality. Since the survey was anonymous, student responses cannot be linked to demographic records. Still, students did answer questions about race, gender, and expected grade that can be used to obtain some preliminary indications of whether further investigation would be warranted.

The first analysis looks at whether expected grade made a difference in student ratings. Students were asked the question "What grade do you think you will get in this class" with answer choices of A, B, C, D, and F.

The table below displays the results of a regression of a teacher's total score on dummy variables for each level of a student's expected grade with the expected grade of C being the left out group. These results were similar when the same model included controls for student gender and race.

Table 32 - Regression Results from Expected Grade on Student Ratings

Expected Grade	Number of Students	Coefficient	T Statistic
A	4,920	6.93*** (.706)	9.83
B	4,326	3.01*** (.717)	4.19
C	1,457	N/A	N/A
D	187	-1.54 (1.84)	-0.84
F	133	-5.89*** (2.15)	-2.75

* p < .10, ** p < .05, *** p < .001

For a teacher's overall score, there was a strong relationship between the expectation of a higher grade and a student's ratings of this teacher, particularly when a student expected either an A or an F. There are two potential explanations. It is possible that students who expect higher grades rate teachers higher or that students with higher expected grades actually have teachers who more frequently engage in these behaviors. It is likely that a combination of both drives these results. Also, since there are very few students who expected to receive a D or an F it is difficult to make a strong assertion about these students.

The next analysis looks at whether student gender or race influenced ratings. The first model includes controls for being African American, Hispanic, and female. It appears that both Hispanic and African American students tend to rate teachers higher than other students (the omitted group in this case is white male students). When a dummy variable for having a high grade (either an A or B) is included, the coefficient for African American students and Hispanic students are both significant and positive. It also now appears that females have lower ratings when controlling for having a high grade and race. As an added check, model 3 includes an interaction term between having a high expected grade and being African American. The coefficient on African American is no longer significant, but it appears that having a high grade and being African American has a joint impact on student ratings. Overall, it does appear that student characteristics influence student ratings and whether these should be controlled for should be investigated in future work on student surveys.

Table 33 - Regression Results from Demographic Characteristics on Student Ratings

	Model 1	Model 2	Model 3
Black	3.93*** (.486)	3.86*** (.484)	1.52 (1.22)
Hispanic	1.54 (.980)	2.04** (.978)	1.96** (.979)
Female	-.712 (.454)	-1.03** (.453)	-1.02** (.453)
High Grade	X	5.70*** (.617)	4.82*** (.747)
Interaction Term Between Black and High Grade	X	X	2.76** (1.32)

* p < .10, ** p < .05, *** p < .001

Analysis of Missing Data

During the course of survey administration, not all teachers and students that were randomly selected ended up participating. Although the survey did not have high stakes attached to the results (teachers knew that results would not be shared with administrators), it is possible that certain types of teachers did not participate. This could lead to selection bias, thus casting doubt that the relationships found above would hold for all teachers. The section uses limited available data to investigate whether teachers that did or did not participate in the survey had differences in the types of classes they taught.

Although an analysis of student participation patterns could possibly detect whether teachers attempted to influence results by manipulating which students took the survey, it is impossible to know which students did or did not take the survey due to the survey being anonymous for students. Still, it is reasonable to assume that teachers would not have motivation to systematically partake in this behavior due to the clear message sent to teachers that results would not be shared with administrators or Race to the Top staff members. For teachers, however, it is possible that certain types of teachers that were selected may choose not to participate for a variety of reasons including being busy, not wanting to miss class time, fear of survey results, etc.

A total of 835 teachers were randomly selected using methods described earlier. Of these 835, 676 teachers had students participate in the survey. The table below shows the number of teachers that did not participate within each district. The lowest percentage of teachers participating comes from DeKalb County at 59%. DeKalb County was the first district to begin participation and also experienced some technical difficulties when close to 3000 students attempted the survey on the same day. While the research team was able to switch to an unlimited capacity server within 24 hours, survey participants on that day may or may not have

had students retake the survey afterward. This issue also affected teachers within Meriwether County. It is reasonable to assume that teachers participating on this day were not systematically different, but it is possible that teachers who persevered and had students retake the survey may have different characteristics. When results of the survey are rerun using only the remaining five districts, the overall results are similar. Further, when a control is added for testing on this day the results do not change.

Table 34 - Number of Teachers Participating by District

District	# of Teachers Participating	# of Teachers Selected	% of Teachers
DeKalb	121	205	59.0%
Griffin-Spalding	78	78	100%
Hall	167	180	92.8%
Meriwether	67	89	75.3%
Pulaski	39	43	90.7%
Rabun	69	76	90.8%
Savannah	135	164	82.3%

The only other information available on teachers was the name of the course they taught. The table below shows the breakdown of selected teachers that did and did not participate based on the category of courses they taught, with no large differences appearing between the two groups. While the available data are limited, the high degree of participation outside of technical difficulties, the lack of high stakes, and the similarity based on available data provides some evidence that results would be similar if all teachers had participated.

Table 35 - Comparison of Selected and Participating Teachers by Subject

	Math	ELA	Science	Social Studies	Foreign Language	PE/ Health	Elective/ Other	Total
Selected and Participated	128 (20%)	115 (18%)	89 (14%)	70 (11%)	34 (5%)	41 (6%)	159 (25%)	636
Selected but did not Participate	28 (18%)	38 (24%)	29 (18%)	20 (13%)	4 (3%)	11 (7%)	29 (18%)	159

Teacher Survey on Feedback Reports

In the interest of improving the student survey, a teacher response survey was distributed to all participating teachers after they received their feedback reports. An example of this feedback report can be found in Appendix E while the interview questions are included in Appendix F. Teachers were asked about a variety of topics including how accurate they felt the results were, what they found most and least helpful about the results, and whether or not results will influence their classroom practice in the coming year.

A total of 96 out of a possible 667 (14%) teachers responded to the survey. Since the survey was given over the summer, it is possible that many teachers were not regularly checking their work emails or chose not to respond and the sample cannot be considered representative. Still, there is some value is hearing the ways in which these teachers viewed the feedback. Responses have yielded several interesting findings regarding how teachers intend to use results, teachers' perceptions of the student survey, and ways that the student survey could be improved to be more accessible to students and more reflective of teacher practice. Each of these topics is discussed below.

Teachers were asked to describe whether or not the student survey results would influence their teaching in the coming year. Nearly 8 in 10 teachers indicated that the results would change their practice. Planned changes included being more mindful of student needs,

targeting PD toward areas indicated as weaknesses, and incorporating more real-world examples in lessons. Teachers that responded that results would not influence their practice often questioned the accuracy of the survey results or felt that they needed more direction about how to improve their weak areas. Several teacher quotes taken directly from the survey responses are shown below.

> *"Yes, this information will influence my teaching next year. I will be more aware of adjusting my teaching to give students more opportunities for success. The results from this survey will allow me to pick 1 or 2 areas to concentrate on during my teaching, and also give me concrete examples of behaviors which I can ask my co-workers to observe and assist me in improving my teaching practices."*

> *"Not at all. I work hard to address all of the issues mentioned every year and am always looking for ways to improve. Telling me what I need to improve without examples of how to improve in my specific area of foreign language is not beneficial to me in any way."*

> *"Yes, my student feedback has already got me thinking of ways to bridge this gap or disconnect I have with my students. I am looking forward to implementing some new strategies and ideas in my classes."*

For the most part, teachers found the results both helpful and accurate. Teachers were presented with reports that outlined their strengths and weaknesses in each of six performance areas as well as in comparison with the average performance of other teachers in their school and district. Many teachers found the graphic presentation of information helpful. They also appreciated seeing both their strengths and weaknesses. Generally, those that stated they did not find the results helpful questioned the accuracy of the results. Teachers were especially hesitant to trust the sampling design of the survey and said that they would have greater confidence in the results if more of their students had been surveyed. However, over 75 percent of teachers found the student survey results to be very or somewhat accurate.

Based on the 96 responses to the teacher feedback survey, it seems that the large majority of teachers found the survey both helpful and accurate. Nearly 80 percent of teachers indicated that they would use the feedback from their student surveys to influence their classroom practice, and 77 percent found the survey somewhat or very helpful. Teachers said that they intended to use the survey results to guide their PD choices, influence the content and delivery of lessons, and to better serve their students. Respondents also gave a variety of helpful suggestions as to how the survey could be improved such as including a "read-aloud" option and giving more feedback on the performance of teachers in comparison to others in the same subject/grade level.

Chapter 5:

Discussion

The current investigation describes the development and validation of an instrument to measure teacher effectiveness using student feedback. It employs a mixed-method approach to test the survey for its relationship to targeted outcomes as well as internal reliability. Finally, the validity framework includes establishing construct validity through sources of evidence including content validity, convergent validity, and predictive validity. The use of an established validity framework is a unique contribution of the current study, as prior student survey instruments have either not undergone this process or have not documented the results.

Content validity is established through the development of survey questions. Questions ask about behaviors that have been consistently identified in the research as having a positive relationship with academic outcomes. Further, the questions align with validated observation rubrics. Both of these procedures allow for the survey to be both research-based and exhaustive of desired teaching behaviors.

The next aspect of construct validity was investigated through cognitive testing. 25 students and five teachers reviewed survey questions to ensure alignment with objectives as well as readability and comprehension. Cognitive testing was used to determine whether the questions measure what they are intended to measure. Questions were continually revised and retested to reflect the findings from these interviews.

Following the creation and modification of survey questions, pilot testing represented a way of determining convergent and predictive validity. Results a large scale pilot in Georgia demonstrate a positive relationship with all three external measures including value-added

student achievement, academic student engagement, and academic self-efficacy. While correlations with value-added are small and positive, there is a strong relationship between a teacher's total score and measures of academic engagement and self-efficacy. Results for value-added varied by grade level and subject, with the stronger relationships between a teacher's total score in science and social studies than ELA and math. Further, there were stronger relationships with ELA and science for middle school students than high school while the opposite was true for math and social studies. Overall results for all subjects, however, were significant at least at the $p<.1$ level.

In the policy context, there are several important issues to consider when choosing to adopt student surveys. Unfortunately, many of these do not have research available to assist in making an informed choice. First, a decision must be made regarding whether student surveys will serve as a component of a high stakes teacher evaluation or solely as a method of providing feedback to teachers on their instructional practices. Though results provide preliminary evidence that teachers had intentions of incorporating feedback from student surveys, there was no follow-up on whether teachers actually implemented the suggestions or whether these changes had any impact on student outcomes. Further, it is unclear whether using feedback reports in tandem with coaching from lead teachers or principals would better facilitate instructional change.

In a high stakes setting, there are several issues to consider. First, there is no consensus on what percentage of a teacher's evaluation should come from student survey results. The next round of the MET project aim to provide insight on this question, but policy makers must decide whether to give stronger weight to metrics that hold a stronger relationship with desired outcomes, whether to base the percentages on stability of estimates, or whether to develop a strategy that fits within the existing policy context. Next, it is unclear whether student ratings

would be similar in high stakes and non-high stakes context. Future research is outlined below, but it is possible that student ratings may change based on teacher or student characteristics in a high-stakes environment.

On a related topic, it is not clear how teachers' behaviors would change when student surveys count towards their evaluation. It is possible that some teachers would attempt to influence student ratings in both desired as well as unintended ways. Having controls in place (such as questions that ask students directly about teacher attempts to influence ratings) as well as focusing on items that are less responsive to negative teacher influence are potential solutions that have yet to be explored.

There are also issues that pertain to both high stakes and feedback only settings. First is how many classes or students should be used in order to determine a teacher's overall rating. Using more classes has the benefit of somewhat greater accuracy and increased teacher buy-in since teachers could feel it is a more representative sample of their classes. Conversely, using fewer classes could possibly achieve similar results without the disadvantages of missing more class instruction and students growing fatigued after 6-8 surveys. Further, there is no evidence on the number of times a teacher should be rated by their students each year. It is possible that multiple evaluations could provide more reliable estimates and also reflect growth during the year. Finally, there are several potential options for how survey items values lead to a teacher's overall score. Options include weighting certain items, counting all items equally, or giving equal weight to each of the different scales (presenter, manager, etc.).

We are still at a very early stage of using student surveys as a measure of teacher evaluation. Though further investigation into specific details of student surveys is essential, it is important to conduct these studies with a study that possesses strong metrics both internally and externally and has been thoroughly validated. The minimum number of students required to take

the survey, whether answers differ depending on the stakes for the teachers, and whether screening procedures are effective all are relevant questions that can now be better investigated using the instrument developed in the current study.

Recommendations for Student Survey Development and Use in Teacher Evaluation

- Cognitive interviews: While statistical analysis can provide insight into which questions have relationships with desired outcomes, this technique is less adept at determining why a question may not show a strong relationship. For instance, one item that informed the data was "I learn from mistakes in this class". Instead of eliminating this question, it was found through subsequent interviews that students were unaware of whether the question was referring to academic or behavior mistakes. The question still has value if adjusted to reflect the focus on academic mistakes and should then be retested for its relationship to outcome measures.

- Avoid negatively worded questions: Students continually showed a tendency to either misinterpret the question or be less likely to choose the lower end of the scale. While negative questions are important to include as a means of preventing a continual response pattern, these questions should likely not be included in calculating a teacher's overall average.

- Use screening procedures: Although uncommon, there were a number of students that did not answer questions carefully. Primarily this consisted of students responding to all questions with the same answer choice. The elimination of these

responses results in a more accurate evaluation and will provide more helpful feedback to teachers.

- Investigate controlling for student characteristics: The analysis of how student characteristics influence ratings suggests that students with a higher grade expectation rate their teachers more favorably. While it is possible that these students have access to better teachers, it would be important to consider controlling for prior student grades or test scores when calculating teacher averages on student surveys.

- Provide feedback for teachers: Despite the limited number of teachers that responded to the survey, many teachers within the sample reported valuing the feedback provided in the teacher reports. Specifically, teachers identified areas for improvement and suggested work with colleagues on developing effective teaching strategies that met this need. Further, it is possible that teachers will be more invested in using student surveys as a measure of teacher evaluation when they see the teacher reports.

Future Investigations

The use of student surveys as a measure of teacher evaluation is still in the very early stages. As such, there are several unanswered questions that remain regarding the use of student surveys that will aid policy makers in decisions regarding their use as a measure of teacher effectiveness. Several of these areas for future research are described below.

Critics argue that students would be incapable of providing accurate feedback, particularly when the responses are part of a high stakes evaluation for a teacher. As Jesse

Rothstein notes in his review of the findings from the MET project, "Mischievous adolescents given the opportunity to influence their teacher's compensation and careers via their survey responses may not answer honestly… studies of zero stakes student surveys can tell us little about how the students would respond if their teachers' careers was on the line" (Rothstein, 2010, p. 7).

Some of this concern is derived from the broader evaluation literature. In the private sector, for instance, there is some evidence of performance appraisals being influenced by the stakes (Fried, 1999). Further, the human resource literature suggests that raters are more critical when ratings are used for research rather as opposed to administrative practices (Murphy & Cleveland, 1995). Additionally, there has been research in the field of mock juries that suggests that the consequences of the situation may affect the actual judgment. As the authors note, "a participant may make choices other than what he or she would if the study conditions were real, the stakes can matter, and the failure to account for them can be very problematic" (Cahoy & Ding, 2006, p. 1276).

A possible way of providing insight on this concern is by administering a student survey on teacher effectiveness in both high- and low-stakes settings employing a randomized control experimental design in school districts. In order to create a high stakes environment, students receive a survey with instructions that outline how the results will impact the teacher. For the high-stakes condition, the instructions would indicate that the results provide feedback for the teacher and that the results will be part of the teacher's yearly evaluation that determines whether the teacher's contract is renewed. For the low-stakes comparison, the directions would only say that the results will provide feedback for the teacher.

The analysis would compare overall mean survey scores for teachers in high-and low-stakes settings to determine whether a difference exists for overall teacher scores on the survey. Part of the analysis would examine whether the responses varies by age of students. In addition, it would examine how well these evaluations correlate with principal evaluations and value added assessments of teachers.

In education, there is some evidence that teachers respond to high stakes environments by altering their content coverage and assessment methods so that they are aligned with the test (Darling-Hammond& Wise, 1985; Furman, Clune, & Elmore, 1991; Koretz, Barron, Mitchell, & Stecher, 1996; Mehrens, 1998; Rosenholtz, 1987; Rowan, 1996). In addition, some researchers argue that high-stakes testing environments lead to greater student anxiety and disengagement from school (Linn, 1994; Mehrens, 1998; Wheelock, Bebell, & Haney, 2000). For instance, Trippett and Barksdale (2005) used students drawings and written descriptions on the day after a test to analyze the effect of high stakes and low stakes testing for 225 elementary students in 5 different areas. The students were most likely to describe nervousness, isolation, confusion and anger. Together, these studies suggest a plausible hypothesis that students' evaluations may be affected by the stakes associated with the evaluation. However, there is no research documenting whether student ratings in education are influenced by the ways in which the results will ultimately impact the teacher.

Other possible investigations include a more detailed investigation of how teachers incorporate feedback from student surveys. While this analysis provided preliminary information about what teachers thought of the feedback reports, a more systematic study could analyze whether the suggestions for improvement were actually implemented and whether this had an impact on student achievement. Further, it would be possible to incorporate coaching in

similar ways as previous studies of coaching using feedback for principals (Bickman, Goldring, Andrade, Breda, & Goff, 2012).

To properly conduct any of the investigations it is essential to have an instrument that has undergone extensive testing and validation work. The work from the current investigation will provide more confidence in the findings from these studies. Though the use of student surveys is still in its infancy, the potential for use within systems of teacher evaluation becomes more of a possibility with the work outlined above.

WORKS CITED

Alchian, A., & Demsetz, H. (1972). Production, information costs, and economic organization. *American Economic Review*, 62(5): 777-795.

Aleamoni, L. (1999). Student Ratings Myths Versus Research Facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for Student Background in Value-Added Assessment of Teachers. *Journal of Educational Behavior and Statistics*.

Beatty, P., & Willis, G. (2007). Research Synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 287-311.

Bickman, L., Goldring, E., Andrade, A., Breda, C., & Goff, P. (2012). Improving Principal Leadership Through Feedback and Coaching. *SREE.* Washington, DC.

Blair, J., & Brick, P. (2009). *Current Practices in Cognitive Interviewing.* Rockville, MD: Westat.

Brophy, J., & Good, T. (1986). Teacher Behavior and Student Achievement. In M. Wittrock, *Handbook of Research on Teaching* (pp. 340-370). New York: MacMillan.

Burniske, J., & Meibaum, D. (2011). *The Use of Student Perceptual Data as a Measure of Teaching Effectiveness.* Texas Comprehensive Center.

Cahoy, D., & Ding, M. (2006). Stakes Matter: Empirical Evidence of Hypothetical Bias in Case Evaluation and the Curative Power of Economic Incentives. *St. John's Law Review*.

Chetty, R., Friedman, J., & Rockoff, J. (2011). *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood.* Cambridge, MA: NBER.

Clotfelter, J., Ladd, H., & Vigdor, J. (2006). *Teacher-student matching and the assessment of teacher effectiveness.* Cambridge, MA: National Bureau of Economic Research.

Dee, T. (2004). The Race Connection. *Education Next*, 52-59.

Desimone, L. M., & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Education Evaluation and Policy Analysis*, 1-22.

Ehrenberg, R., Goldhaber, D., & Brewer, D. (1995). Do Teachers' Race Ethnicity or Gender Matter? *Industrial and Labor Relations Review*, 547-561.

Eisenhardt, M. K. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 57.

Emmer, E., & Evertson, C. (1981). Synthesis of Research on Classroom Management. *Educational Leadership*, 342-347.

Ferguson, R. (1991). Paying for Public Education. *Harvard Journal of Legislation*, 458-498.

Fredricks, J., & McColskey, W. (2011). *Measuring student engagement in upper elementary through high school: a description of 21 instruments.* Washington, DC: IES.

Gage, N., & Needles, M. (1989). Process-Product Research on Teaching. *Elementary School Journal*, 253-300.

Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., & Whitehurst, G. (2010). *Evaluating Teachers: The Important Role of Value-Added.* Washington, DC: The Brookings Institution.

Goe, L. (2007). *The Link Between Teacher Quality and Student Outcomes: A Research Synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis.* Washington, DC: National Comprehensive Center on Teacher Quality.

Goldhaber, D. (2002, Spring). The Mystery of Good Teaching. *Education Next*.

Goldhaber, D., & Anthony, E. (2004). *Can teacher quality be effectively assessed?* Washington, DC: Center for Reinventing Public Education.

Goldhaber, D., Goldschmidt, P., Sylling, P., & Tseng, F. (2011). *Teacher Value-Added at the High School Level: Different Model, Different Answers?* Center for Education Data and Research.

Greenwald, R., Hedges, L., & Laine, R. (1996). The effect of school resources on student achievement. *Review of Education Research*, 361-396.

Guarino, C., Reckase, M., & Wooldridge, J. (2011). *Can Value-Added Measures of Teacher Performance be Trusted?* East Lancing, MI: Education Policy Center at Michigan State University.

Hanushek, E. (1992). The Tradeoff Between Child Quantity and Quality. *Journal of Political Economy*, 84-117.

Hanushek, E., Kain, J., O'Brien, D., & Rivkin, S. (2005). *The Market for Teacher Quality.* Cambridge, MA: National Bureau for Economic Research.

Harris, D., & Sass, T. (2007). *Teacher training, teacher quality, and student achievement.* Washington, DC: National Center for the Analysis of Longitudinal Data.

Hill, H., Rowan, B., & Ball, D. (2005). Effect of Teachers' Mathematical Knowledge for Teaching and Student Achievement. *American Education Research Journal*, 371.

Jacob, B., & Lefgren, L. (2005). *Principals as Agents: Subjective Performance Measurement in Education.* Cambridge, MA: Faculty Research Working Paper Series.

Kane, T., & Cantrell, S. (2010). *Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project.* Bill and Melinda Gates Foundation.

Kane, T., & Staiger, D. (2012). *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains.* Seattle, WA: Bill & Melinda Gates Foundation.

Kane, T., Taylor, E., Tyler, J., & Wooten, A. (2010). *Identifying Effective Classroom Practices Using Student Achievement.* Cambridge, MA: NBER.

Kyriakides, L. (2005). Drawing from teacher effectiveness research to establish a teacher evaluation system. *Journal of Classroom Interaction*, 44-66.

Lopez, S. (2009). *Gallup Student Poll National Report.* Gallup, Inc.

MacCaffrey, D., J.R., L., Koretz, D., & Hamilton, L. (2003). *Evaluating Value-Added Models for Teacher Accountability.* Santa Monica, CA: RAND Coorporation.

Marks, H. (2000). Student Engagement in Instructional Activity. *American Educational Research Journal*, 153-184.

Marzano, R., Pickering, D., & Pollock, J. (2001). *Classroom Instruction that Works: Research-Based Strategies for Increasing Student Achievement.* Alexandria, VA: Association for Supervision and Curriculum Development .

McCaffrey, D., Han, B., & Lockwood, J. (2009). Turning Student Test Scores into Teacher Compensation Systems. In M. Springer, *Performance Incentives.* Washington, DC: Brookings Institution Press.

Messick, S. (1989). Validity. In R. Linn, *Educational Measurement.* New York: American Councel on Education/Macmillan.

Moss, P. (2007). Reconstructing Validity. *Educational Researcher*.

Murname, R., & Willett, J. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science Research.* New York: Oxford University Press.

Murphy, K. R., & Cleveland, J. N. (1995). *Performance Appraisal: An Organizational Perspective.* Thousand Oaks, CA: Sage Publications.

Peterson, K., Wahlquist, C., & Bone, K. (2000). Student Surveys for School Teacher Evaluation. *Journal of Personnel Evaluation in Education*.

Pianta, Paro, L., & Hamre. (2006). *Classroom Assessment Scoring System: Preschool version.* Charlottesville, VA: Center for Advanced Study of Teaching and Learning .

Podgursky, M., & Springer, M. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*.

Porter, A., Polikoff, M., Goldring, E., Murphy, J., Elliott, S., & May, H. (2010). Developing a Psychometrically Sound Assessment of School Leadership: The VAL-ED as a Case Study. *Educational Administration Quarterly*, 135-173.

Preacher, K. (n.d.). Psychology 319: Factor Analysis Class Notes. *2012*.

Renaud, R., & Murray, H. (2005). Factorial Validity of Student Ratings of Instruction. *Research in Higher Education*.

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 417-458.

Rockoff, J. (2004). The Impact of Individual Teachers on Student Achievement. *American Economic Review*, 247-252.

Rockoff, J., Jacob, B., Kane, T., & Staiger, D. (2008). *Can You Recognize and Effective Teacher When You Recruit One?* NBER.

Rothstein, J. (2010). *Review of Learning About Teaching.* Boulder, CO: Great Lakes Center for Education Research and Practice.

Rowan, B., Jacob, R., & Correnti, R. (2009). Using Instructional Logs to Identify Quality in Educational Settings. *New Directions for Youth Development*, 13-31.

Rowen, B., Correnti, R., & Miller, R. (2002). What large-scale survey research tells us about teacher effects on student achievement. *Teachers College Record*, 1525-1567.

Sanders, W., & Rivers, J. (1996). *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement.* Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Schacter, J., & Thum, Y. (2004). Paying for High and Low-Quality Teaching. *Economics of Education Review*, 411-430.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling., D. (2009). *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness.* Brooklyn, NY: The New Teacher Project.

Wilkerson, D., Manatt, R., Rogers, M. A., & Maughan, R. (2000). Validation of Student, Principal, and Self-Ratings in 360 Feedback for Teacher Evaluation. *Journal of Personnel Evaluation in Education*.

Appendix A – Research Based Teaching Practices

Rosenshine (1979)	Schachter and Thum (2004) – Teaching Behaviors	Schachter and Thum (2004) – Teaching Strategies	Good and Brophy (1986)	Emmer and Evertson (1994)	Marzano (2001)
Clarity of teacher's presentation and ability to organize classroom activities	Questions – Type, frequency, required student response, wait time	Grouping – strategies for cooperative learning	Clarity about instructional goals Providing students with the opportunity to learn what is to be tested	Rules and Procedures – established and enforced and students are monitored for compliance	Identifying similarities and differences
Variability of media, materials, and activities used by the teacher	Feedback – Frequent, elaborate, and high quality academic feedback	Thinking – Metacognition generative learning	Knowledge of content and ways for teaching it Variety in the use of teacher methods and media Realistic praise – not praise for its own sake	Consistency – Similar expectations are maintained by activities and behavior at all times for all students	Summarizing and Note Taking
Enthusiasm, defined in terms of the teacher's movement, voice inflection, and the like	Lesson Structure and Pacing – Optimizing instructional time	Activities – Meaningful projects and simulations to foster opportunities for learning by doing and student interaction	Making comments that help structure learning of knowledge and concepts for students, helping students learn how to learn	Prompt Management of inappropriate behavior Academic instruction – Attention is focused on the management of student work	Reinforcing Effort and Providing Recognition
Task Orientation or businesslike teacher behaviors, structures, routines, and academic focus	Lesson Objectives-Objectives explicitly communicated	Motivating students – Attend to students notions of competence, reinforcing student effort	With-it-ness – awareness of what is going on, alertness in monitoring classroom activities Overlapping – sustaining an activity while doing something else at the same time	Checking student work – All student work, including seatwork, homework, and papers, is corrected, errors are discussed, and feedback is provided promptly	Homework and Practice
Student Opportunity to Learn, that is, the teacher's coverage of the material or content in class on which students are later tested	Presentation – Illustrations, analogies, modeling by teacher, concise communication	Teacher Knowledge of Students – prior knowledge, incorporating student interest through differentiated approaches	Monitoring of students' understanding, providing appropriate feedback, giving praise, asking questions	Interaction teaching – Presenting and explaining new material, question sessions, discussions, checking for student understanding, actively moving among students, and providing feedback	Nonlinguistic representa-tions
"Promising" -Using student ideas -Justified criticism -Using structuring comments	Classroom Environment – Student discipline and behavior, student work ethic, teacher caring for individual pupils		Smoothness – Sustaining proper lesson pacing and group momentum, not dwelling on minor points or wasting time dealing with individuals, and focusing on all students Flexibility in planning and adapting classroom activities	Clarity – Lessons are presented logically and sequentially. Clarity is enhanced by the use of instructional objectives and adequate illustrations and by keeping in touch with students	Cooperative Learning Questions, Cue, and advance organizers
-Encouraging student elaboration -Using challenging instructional materials			Seatwork instructions and management that initiate and focus on productive task engagement	Pacing – Information is presented at a rate appropriate to the students' ability to comprehend it	Setting Objectives and Providing Feedback

-Asking appropriate questions suited to students' cognitive level			Holding students accountable for learning; accepting responsibility for student learning	Transitions – Transitions from one activity to another are made rapidly, with minimal confusion	Generating and testing hypothesis

Appendix B – Example Coding Scheme for Literature Review

Instructional Goals/Objectives - _____

Asking Questions - _____

Presentation of Material - _____

Providing Feedback - _____

Reinforcement/Praise - _____

Classroom Environment - _____

Rosenshine (1979)	Schachter and Thum (2004) – Teaching Behaviors	Schachter and Thum (2004) – Teaching Strategies	Good and Brophy (1986)	Emmer and Evertson (1994)	Marzano (2001)
Clarity of teacher's presentation and ability to organize classroom activities	Questions – Type, frequency, required student response, wait time	Grouping – strategies for cooperative learning	Clarity about instructional goals	Rules and Procedures – established and enforced and students are monitored for compliance	Identifying similarities and differences
Variability of media, materials, and activities used by the teacher**	Feedback – Frequent, elaborate, and high quality academic feedback	Thinking – Metacognition generative learning	Knowledge of content and ways for teaching it	Consistency – Similar expectations are maintained by activities and behavior at all times for all students	Summarizing and Note Taking
Enthusiasm, defined in terms of the teacher's movement, voice inflection, and the like**	Lesson Structure and Pacing – Optimizing instructional time	Activities – Meaningful projects and simulations to foster opportunities for learning by doing and student interaction	Variety in the use of teacher methods and media	Prompt Management of inappropriate behavior	Reinforcing Effort and Providing Recognition
Task Orientation or businesslike teacher	Lesson Objectives- Objectives explicitly	Motivating students – Attend to students notions of competence,	With-it-ness – awareness of what is going on, alertness in	Checking student work – All student work, including	Homework and Practice

behaviors, structures, routines, and academic focus	communicated	reinforcing student effort	monitoring classroom activities	seatwork, homework, and papers, is corrected, errors are discussed, and feedback is provided promptly	
Student Opportunity to Learn, that is, the teacher's coverage of the material or content in class on which students are later tested	Presentation – Illustrations, analogies, modeling by teacher, concise communication	Teacher Knowledge of Students – prior knowledge, incorporating student interest through differentiated approaches	Overlapping – sustaining an activity while doing something else at the same time	Interaction teaching – Presenting and explaining new material, question sessions, discussions, checking for student understanding, actively moving among students, and providing feedback	Nonlinguistic representa-tions
"Promising" -Using student ideas -Justified criticism -Using structuring comments	Classroom Environment – Student discipline and behavior, student work ethic, teacher caring for individual pupils		Smoothness – Sustaining proper lesson pacing and group momentum, not dwelling on minor points or wasting time dealing with individuals, and focusing on all students	Academic instruction – Attention is focused on the management of student work	Cooperative Learning
-Encouraging student elaboration -Using challenging instructional materials			Seatwork instructions and management that initiate and focus on productive task engagement	Pacing – Information is presented at a rate appropriate to the students' ability to comprehend it	Setting Objectives and Providing Feedback

-Asking appropriate questions suited to students' cognitive level			Holding students accountable for learning; accepting responsibility for student learning	Transitions – Transitions from one activity to another are made rapidly, with minimal confusion	Generating and testing hypothesis

Appendix C – Questions Organized according to Danielson Framework

(Note: Planning and Preparation and Professional Responsibilities are not included)

Classroom Environment

-Creating an environment of respect and rapport: Interactions among teacher and individual students are highly respectful, they reflect genuine warmth and caring, sensitivity to students' backgrounds and levels of development, students themselves ensure high levels of civility among members

Survey Question (CE3): My teacher shows respect for all students.

-Establishing a culture for learning: High levels of student engagement and teacher passion for the subject create a culture for learning, everyone shares the belief that the subject is important, all students hold themselves to a high standard of performance, teacher and students demonstrate high level of respect for knowledge of diverse student cultures

Survey Question (CK2a): My teacher is enthusiastic about the subject

Survey Question (M2): My teacher helps me understand why the things we're learning in class are important to know in life.

Survey Question (CE4): My teacher expects me to take pride in the quality of my work for this class.

-Managing classroom procedures: Students contribute to the seamless operations of classroom routines and procedures

Survey Question (LS3): Students help the teacher with classroom tasks (passing out papers, materials, etc.)

-Managing student behavior: Standards of conduct are clear, with evidence of student participation in setting them, teacher's monitoring of behavior is subtle and preventive, teacher's response to student misbehavior is sensitive to individual student needs, students take an active role in monitoring the standards of behavior

CLASS Survey Question: My teacher explains how we are supposed to behave in class.

CLASS Survey Question: I understand the rules for behavior in this class.

CLASS Survey Question: My teacher walks around the room to check on students when we are doing individual work in class

New Survey Question: The students help to come up with the rules for the class (Check that this makes sense as a frequency question)

-Organizing physical space: The classroom is safe, technology is used skillfully as appropriate to the lesson

New Survey Question: My teacher uses technology (computers, sensors, videos, etc) in class.

Instruction

-Communicating with students: Expectations for learning, directions and procedures, and explanations of content are clear to students. Teacher's oral and written communication is clear and expressive, appropriate to students' diverse cultures and levels of development, and anticipates possible student misconceptions

Survey Question (P1): My teacher explains information in a way that makes it easier for me to understand.

Survey Question (P3): When explaining new skills or ideas in class, my teacher tells us about mistakes that student might make.

-Using questioning and discussion techniques: Questions reflect high expectations and are culturally and developmentally appropriate. Students formulate many of the high-level questions and ensure that all voices are heard.

Survey Question (Q1): My teacher asks questions in class that make me really think about the information we are learning

Survey Question (Q2a): When my teacher asks questions, he/she only calls on students that volunteer (reverse)

Survey Question (Q2b): When my teacher asks questions, he/she calls on all students equally (boys, girls, etc.)

New Survey Question: Students ask challenging questions during class.

-Engaging students in learning: Students are highly intellectually engaged throughout the lesson in higher order learning, and make material contributions to the activities, student groupings, and materials. The lesson is adapted as needed to the needs of the individuals, and the structure and pacing allow for student reflection and closure. Students assist in ensuring that activities, assignments and materials are fully appropriate for diverse cultures.

Survey Question (LS2): At the end of each lesson, the teacher has us summarize or talk about what we have just learned.

Survey Question (LS1): We are learning or working during the entire class period.

Survey Question (G1): When working in groups, my teacher has us choose a job, role, or responsibility within the group (recorder, materials person, etc)

Survey Question (A2): The activities we do in class keep me interested.

New Survey Question: This class is challenging.

-Using assessment in instruction: Multiple assessments are used in instruction, through students involvement in establishing assessment criteria, self-assessment by students and monitoring of progress by both students and teachers, and high quality of students from a variety of sources

Survey Question (F1a): My teacher provides written comments on assignments.

Survey Question (F3): My teacher checks to see if I understand what we're learning during the lesson.

Survey Question (F4): I have opportunities to give and receive feedback from other students in the class.

Survey Question (F2): My teacher gives us guidelines for assignments (rubrics, charts, grading rules, etc) so we know how we will be graded.

New Survey Question: My teacher allows students to help set guidelines for assignments.

New Survey Question: My teacher gives me opportunities to show what I know in different ways (tests, projects, presentations, etc).

-Demonstrating flexibility and responsiveness: Teacher is highly responsive to individual students' needs, interests and questions, make even major lesson adjustments as necessary to meet instructional goals, and persists in ensuring the success of all students.

Survey Question (P2b): If I do not understand something in class, my teacher explains it in a different way to help me understand.

New Survey Question: My teacher is not satisfied until all students understand what we are learning.

New Survey Question: My teacher changes the activity or lesson if many students do not understand.

Survey Question (TS2): My teacher encourages us to ask questions in class.

Appendix D - Questions Organized according to CLASS

Emotional Support

Positive Climate

-Relationships: Teachers and students enjoy interactions with each other, they are interested in spending time with each other, they have an interest in each other's lives outside of school

New Survey Question: My teacher is interested in my life outside of school.

Survey Question (TS3): My teacher cares how I do in school.

-Positive Affect: Teachers and students are smiling and laughing, enjoyment and positive energy, students and teacher appear to be enthusiastic and to enjoy class activities

New Survey Question: I look forward to coming to this class.

New Survey Question: My teacher seems to enjoy teaching this class.

-Positive Communications: Teacher shares positive comments with students, teacher communicates positive expectations for students

Survey Question (M1b): My teacher believes that I can do well in this class.

New Survey Question: My teacher tells me when I do something well.

-Respect: Language that communicates respect, students and teachers have calm and warm voices when speaking to one another, students are cooperative with each other

Survey Question (CE3): My teacher shows respect for all students.

Negative Climate

-Negative Affect: Teachers and/or students are irritated by each other, use harsh voices with each other, engage in aggressive acts, the teacher and/or students frequently express annoyance, irritation or anger without a clear reason, irritation escalates

New Survey Question: My teacher gets angry with students during class.

-Punitive Control: Teacher yells, threatens to punish, or actually punishes students that misbehave. Teacher engages in physical controls such as pushing or pulling students to respond.

New Survey Question: My teacher threatens to punish us.

New Survey Question: My teacher yells at us during class.

-Disrespect: Pattern of disrespect through teasing, bullying, humiliation, or sarcasm, language or behavior that is inflammatory (reference to drugs, sex, alcohol), discriminatory (racism, sexism, or sexual harassment), or derogatory (belittling, degrading)

New Survey Question: My teacher says mean things to students in class.

Teacher Sensitivity

-Awareness: Checks in with students, anticipates problems, notices when a student is struggling to understand or appears upset, notices when students are not engaged in a task

Survey Question (F3): My teacher checks to see if I understand what we're learning during the lesson.

Survey Question (P3): When explaining new skills or ideas in class, my teacher tells us about mistakes that student might make.

New Survey Question: My teacher notices when I am not participating in class.

-Responsiveness to academic and social/emotional needs and cues: Teacher responds to struggling student by providing direction, assistance, and reassurance, adjusts pacing according to what students need, reengages students that are not fully participating, considers outside factors as needed, responds to students who have their hand raised

New Survey Question: If many students do not understand something during the lesson, my teacher changes the way he/she is teaching that idea.

New Survey Question: My teacher calls on students when they raise their hand to ask a question.

Survey Question (Q2a): When my teacher asks questions, he/she only calls on students that volunteer (reverse)

-Effectiveness in addressing problems: Students seemed to be helped after interactions, teacher follows up with students that had difficulty

Survey Question (P2b): If I do not understand something in class, my teacher explains it in a different way to help me understand.

New Survey Question: If I do not understand something in class, my teacher works with me until I understand.

-Student comfort: Students seek out the teacher for assistance, teacher allows students to take risks, students freely share their ideas and attempt to answer difficult questions

Survey Question (TS2): My teacher encourages us to ask questions in class.

New Survey Question: I feel comfortable trying to answer a question in class even if I'm not sure that I am right.

Regard for Adolescent Perspectives

-Support for student autonomy & leadership: Students have choice in assignment, students have responsibility within the classroom, have opportunities to assume responsibility for their own learning

Survey Question (M3): My teacher gives me opportunities to investigate the parts of the subject that interest me the most.

Survey Question (LS3): Students help the teacher with classroom tasks (passing out papers, materials, etc.)

-Connections to current life: Connect content to students' experiences or to current adolescent culture, consistently explains the usefulness of mastering content or skills, students understand why the information or skills presented are important

Survey Question (M2): My teacher helps me understand why the things we're learning in class are important to know in life.

New Survey Question: Possible question on using outside culture? ***

-Student ideas and opinions: Activities and lessons provide opportunities for students to share their ideas, teacher is flexible and attentive to student responses and uses these responses in the lesson

New Survey Question: My teacher encourages me to share my ideas or opinions about what we are learning in class.

-Meaningful peer interactions: Lessons or activities promote constructive peer interactions, students talk openly with each other in a free exchange

New Survey Question: I have opportunities during this class to discuss what we are learning with my classmates during class.

-Flexibility: Teacher provides student freedom of movement.

Classroom Organization

Behavior Management

-Clear expectations: Rules and expectations for behavior are clearly stated and/or understood by all members of the class. Enforced in a consistent and predictable manner. May or may not review expectations. No confusion by students regarding rules and behavioral expectations.

New Survey Question: My teacher explains how we are supposed to behave in class.

New Survey Question: My teacher corrects students when they do not follow the rules of the class.

New Survey Question: I understand the rules for behavior in this class. (I understand how I am supposed to behave/act in this class)

-Proactive: Teacher monitors the classroom, proactive instead of reactive discipline, teacher walks around the room during individual work to reinforce students' on-task behavior, uses proximity and notes positive examples of behavior

New Survey Question: My teacher walks around the room to check on students when we are doing individual work in class.

New Survey Question: My teacher tells us when we are behaving well.

-Effective redirection of misbehavior: Effective subtle means of redirecting students, teacher encourages students to settle disputes on their own first, problems are resolved quickly and effectively, very little time actually managing behavioral problems

New Survey Question: My teacher spends a lot of time in class dealing with poor student behavior (reverse)

-Student misbehavior: Students meet expectations for behavior without many reminders

Survey Question (CE1): Our class is interrupted because of poor student behavior (reverse).

New Survey Question: Students sleep during class (reverse)

Productivity

-Maximizing learning time: Time for learning is maximized, clear directions/options for students that finish early, don't have to be engaged but should be doing something, teacher is fully prepared for lessons and materials are ready and easily accessible, minimizing the number and length of disruptions to learning

Survey Question (LS1): We are learning or working during the entire class period.

New Survey Question: My teacher has something for me to do if I finish an in-class assignment early.

New Survey Question: We spend time in class waiting for the teacher to get everything ready for the next activity. (reverse)

-Routines: Students know what they should be doing. Students show little confusion about routines. "well-oiled" machine where everybody knows what is expected of them.

-Transitions: Little wasted time as student move from one activity to the next. Students are redirected to the next task quickly.

Instructional Learning Formats

-Learning targets/organization: Clearly communicates learning objectives, students appear aware of the point of the lesson, previewing or advance organizers, clear summaries are provided, information presented is well organized and accessible to students

Survey Question (LO1): My teacher tells us about the learning goals/objectives of the day.

Survey Question (LS2): At the end of each lesson, the teacher has us summarize or talk about what we have just learned.

-Variety of modalities, strategies, and materials: Teacher uses different modalities and strategies in order to present information in many ways. Students become actively engaged through manipulating and exploring the resources. Limited use of lecture that has no student participation, oral explanations are reinforced by interesting visuals.

New Survey Question: We learn in many different ways during class (lecture, working in groups, projects, student presentations, etc.).

-Active facilitation: Active facilitator or student participation by asking students questions, lessons are appropriately paced so students are consistently involved, teacher conveys interest in the subject through facial expression, tone, etc.

Survey Question (LS4): The teacher presents material at a speed that I can understand.

Survey Question (CK2a): My teacher is enthusiastic about the subject.

-Effective engagement: Students are focused on important work. Listening to the teacher, raising their hands or volunteering information, actively participating in discussions, group, or individual work

Instructional Support

Content Understanding

-Depth of Understanding: Students apply their thinking to real world situations, teacher presents multiple points of view or perspectives, students should understand different perspectives and not just the opinion of the teacher, student practice new procedures and skills

Survey Question (ST2): My teacher has me apply what we are learning to real-life situations.

New Survey Question: I have a chance to practice new skills or procedures that we learn in class.

-Communication of concepts and procedures: Teacher defines the essential characteristics of the content or procedures, presents multiple and varied examples and non-examples, conditions or appropriate use for procedures

Survey Question (P2a): My teacher uses examples or illustrations to help explain ideas.

-Background knowledge and misconceptions: New information is linked to background information, integrates new information into existing framework, clarifies misconceptions, encourages students to share knowledge and make connections

Survey Question (LO2): My teacher explains how new ideas relates to what we have previously learned.

-Transmission of content knowledge and procedures: Clear and accurate definitions of content are provided, teacher can answer students' questions

Survey Question (P1): My teacher explains information in a way that makes it easier for me to understand.

Survey Question (CK1): My teacher is able to answer students' questions about the subject.

Analysis & Problem Solving

-Opportunities for higher level thinking: Teacher promotes student use of higher level thinking by providing challenging activities or questions. Analysis – separate concepts into parts so that its organizational structure can be understood, Creation/synthesis – put together parts to form a whole with emphasis on creating a new meaning or structure, Evaluation – student make judgments about the value of ideas. Provides structure and time for students to think independently with questions that require divergent thinking.

Survey Question (Q1): My teacher asks questions in class that make me really think about the information we are learning

-Problem solving: Students are challenged to identify the problem, apply existing knowledge to new applications in order to solve the problem. Teacher facilitates students' problem solving techniques instead of showing them how to do it.

New Survey Question: My teacher has me use what I am learning about to solve new problems.

-Metacognition: Thinking out loud, student should reflect on their thought processes, students evaluate their own work, teacher models the thinking out loud process

New Survey Question: My teacher asks me to think about how I come up with my answers.

Quality of Feedback

-Feedback loops: Multiple instances when teachers and students engage in back and forth exchanges, feedback among peers, sustained interaction or persistence in the feedback process

Survey Question (F4): I have opportunities during this class to give and receive feedback from other students.

Survey Question (F1a): My teacher provides written comments on assignments.

-Prompting thought processes: Students are asked to explain their thinking and rationale for responses and actions, extend responses when they give a correct answer or when they give an incorrect answer

New Survey Question: When I answer a question wrong in class, my teacher helps me figure out the right answer.

New Survey Question: When I say the right answer in class, my teacher asks me to explain how I came up with my answer.

-Scaffolding: Teacher provides students with assistance and hints that help students perform academic tasks, teacher prompts students to help scaffold, when student is struggling the teacher provides help rather than moving on

New Survey Question: If I make a mistake, my teacher gives me hints that help me figure out what I did wrong.

-Providing information: Teacher expands student responses in order to provide more information of clarification, teacher gives specific feedback that is individualized to students or contexts

-Encouragement and affirmation: Teacher offers encouragement of student effort that increases involvement and persistence, teacher focuses attention on effort

Survey Question (M1a): My teacher helps me believe that working hard in this class will benefit me.

Student Outcome

Student Engagement

-Active engagement: Student are actively engaged in classroom discussion and activities, asking their own questions, appear to be on task and focused on class-related goals, sharing ideas

New Survey Question: My teacher encourages me to participate in class discussions.

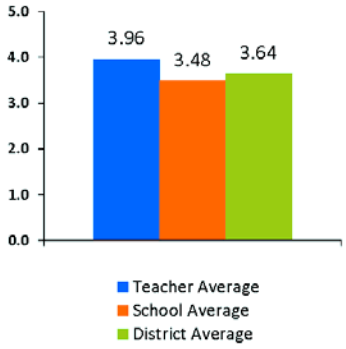
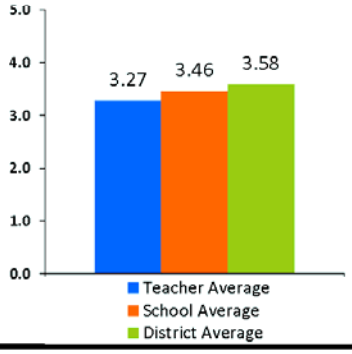
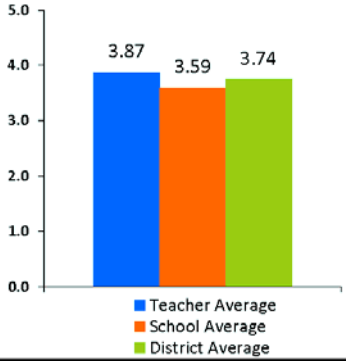
-Sustained engagement: Engagement is sustained through different activities and lessons, student appear interested in and involved in the activities that the teacher has planned

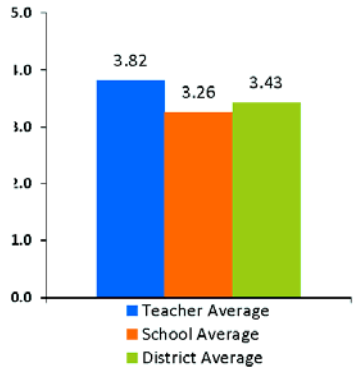
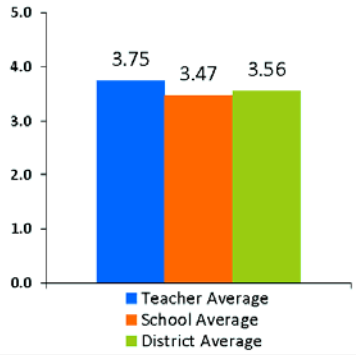
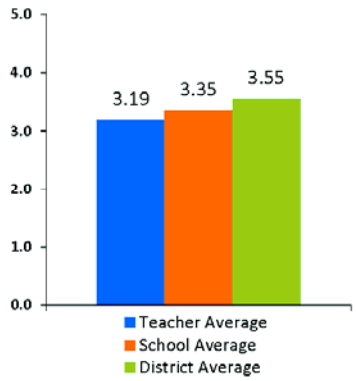
Survey Question (A2): The activities we do in class keep me interested.

Appendix E – Sample Teacher Report

2010-11 Student Survey Results

Teacher: Sample Teacher School: Georgia High School District: Georgia

Teaching Skill Category (1-5 scales, 5=strong performance)	Strengths and Areas for Improvement
Presenter Ability to present information 	**Presentation Style** Strengths: Giving examples of common student mistakes (4.41) Areas for Improvement: Explaining how new ideas are connected to previously learned concepts (3.53)
	Lesson Structure Strengths: Going through examples together with the class (4.65) Areas for Improvement: Using technology in class that helps students learn (computers, sensors, videos, etc.) (3.07)
Manager Ability to manage a classroom 	**Behavior Management** Strengths: Enforcing class rules (4.08) Areas for Improvement: Telling students when they are behaving well (3.02)
	Time Management Strengths: Minimizing time spent transitioning between activities (4.06) Areas for Improvement: Providing students with additional learning opportunities if they finish classwork early (2.29)
Counselor Ability to support students 	**Teacher-Student Relations** Strengths: Showing respect for all students (4.54) Areas for Improvement: Showing interest in students' academic performance in other classes (2.92)
	Awareness of Student Needs Strengths: Noticing when students are not participating in class (3.82) Areas for Improvement: Being available to help students outside of class (3.17)

Teaching Skill Category (1-5 scales, 5=strong performance)	Strengths and Areas for Improvement
Motivator Ability to engage students in learning 	**Investing Students** Strengths: Explaining why things learned in class are important (4.24) Areas for Improvement: Having students apply what they are learning to real-life situations (3.59) **Engaging Students** Strengths: Encouraging students to share ideas/opinions about what they are learning (4.29) Areas for Improvement: Calling on all students in class, not just those who volunteer (3.64)
Coach Ability to provide feedback and challenge students 	**Providing Feedback** Strengths: Providing grading guidelines for assignments (grading rules, charts, rubrics, etc.) (4.12) Areas for Improvement: Communicating with parents about how students are doing in class (2.92) **Challenging Students** Strengths: Expecting student to do their best on assignments (4.53) Areas for Improvement: Helping students figure out the answer when they give an incorrect response in class (3.18)
Content Expert Knowledge of subject material 	**Content Knowledge** Strengths: Being able to answer student questions about the subject (3.53) Areas for Improvement: Bringing in outside materials about the subject (news articles, real-life examples, etc.) (2.29) **Encouraging Student Thinking** Strengths: Providing students with opportunities to show what they know in different ways (tests, projects, presentations, etc.) (4.18) Areas for Improvement: Asking students to explain the thinking behind answers (2.82)

Number of With Valid Survey Responses: 17

Appendix F – Interview Questions for Teachers on Feedback Report

Did you look through your student survey feedback results?

If you didn't look through your results, why not?

What did you find the most helpful on your teacher feedback report?

What did you find the least helpful on your teacher feedback report?

Will your student survey feedback influence your teaching next year? How? If not, why not?

Did you find the results from your survey helpful?

Did you find your results to be accurate?

What other information would you like to have on your report?

What changes would you make to the student survey that your students took this past spring?

What else would you like to share with the researchers about either the feedback report or the survey?

What is your name? (Not required)

What district do you teach in?

Would you be interested in seeing sample videos of teachers that performed well in each category (presenter, manager, etc.)?